الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

وزارة التعليم العالي و البحث العلمي

Minister of Higher Education and Scientific Research

جامعة عباس لغرور – خنشلة

Abbes Laghrour University Khenchela, Algeria

Faculty of Science and Technology
Department of Electrical Engineering

THESIS

Submitted in Candidacy for the Degree of Doctorate in Third Cycle

Speciality: Automation & Industrial Informatics

Design And Development of an Intelligent Fire Risk Prediction System Based on Internet of Things (IOT) Technologies

---

Presented by:

Mohamed LAMRI

Before the Jury composed of:

Dr. Sofiane	**BOUOUDEN**	**Professor**	**Univ. of Khenchela**	**President**
Dr. Abdelhakim	**SAHOUR**	**MCA**	**Univ. of Khenchela**	**Supervisor**
Dr. Farouk	**BOUMEHREZ**	**MCA**	**Univ. of 8 Mai 45-Guelma**	**Co-supervisor**
Dr. Fouad	**ALOUANI**	**Professor**	**Univ. of Khenchela**	**Examiner**
Dr. Salah	**CHENIKHER**	**Professor**	**Univ. of Tebessa**	**Examiner**
Dr. Hanane	**DJELLAB**	**Professor**	**Univ. of Tebessa**	**Examiner**
Dr. Abdelaali	**BAKHOUCHE**	**MCA**	**Univ. of Khenchela**	**Invited**

Academic year: 2025/2026

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

In the name of God, the Most Gracious, the Most Merciful

To the pillars of my life and memory:

To my beloved wife, whose unwavering love, patience, and support
formed the foundation of this journey.
Your strength sustained us through every challenge.

To my son and two precious daughters,
the lights of my life who brought joy and purpose to our home
during the longest nights of writing.

To my father and mother,
whose lifelong sacrifices, boundless love, and steadfast belief
instilled perseverance and wisdom in my soul.

In loving memory of my dear brother ***Said****,*
whose spirit continues to guide and inspire me,
and whose absence reminds me to cherish every moment.

This work is as much yours as it is mine.

Acknowledgements

I express my profound gratitude to the individuals and institutions whose support proved indispensable for completing this doctoral research. First, I extend my deepest appreciation to my doctoral advisors, Professor **Abdelhakim Sahour** and Professor **Farouk Boumehrez**, for their unwavering guidance, intellectual mentorship, and steadfast encouragement throughout this demanding journey. Their insightful critiques and profound expertise were instrumental in shaping this work.

I am also deeply grateful to **Dr. Bakhouche Abdelaali** for his attentive mentorship, steadfast guidance, and unwavering trust throughout my research journey.

My sincere thanks extend to the members of my doctoral committee for their valuable time, critical perspectives, and constructive feedback, which significantly strengthened the thesis.

This research was made possible through the generous institutional support of the **Laboratory of Systems and Applications of Information and Telecommunications Technologies (SATIT)**, the **Department of Electrical Engineering, Abbes Laghrour University, Khenchela**, and funding from Professor **Soufiane Bououdene**. I am indebted to the technical and administrative staff whose efforts facilitated this work.

I extend my profound appreciation to our key socio-economic partners: the **Forest Conservation Departments of Khenchela** and the **Algerian Civil Protection Directorate**. Their collaboration, provision of local expertise, and access to field sites proved crucial for the empirical validation of this trans-disciplinary work. Special recognition is due to the forest rangers and field technicians whose on-ground dedication made this research possible.

My gratitude also extends to our international collaborators, including researchers from **Les Mureaux, France**, particularly Professor **Benkherat**, for fostering a stimulating trans-disciplinary dialogue and providing valuable comparative perspectives.

Finally, to my family and friends—your unconditional support, patience, and understanding during this endeavor have been my greatest source of strength. This achievement is as much yours as it is mine.

الملخص

يشكل تزايد حرائق الغابات عالمياً والمتفاقم بسبب التغير المناخي، تهديداً اجتماعياً وبيئياً خطيراً في المناطق المتوسطية مثل الجزائر. وتثبت أنظمة المراقبة التقليدية التي تعتمد على مكافحة الحرائق برد الفعل والخريطة الثابتة للمخاطر، إلى جانب نماذج الذكاء الاصطناعي غير الشفافة، عدم كفايتها للإدارة الوقائية والموثوقة للمخاطر. يواجه بحث الدكتوراه هذه القيود من خلال تطوير إطار عمل جديد متعدد التخصصات للتنبؤ والتخفيف من مخاطر حرائق الغابات في الوقت الفعلي والقابل للتفسير.

يكمن الابتكار الأساسي في التكامل التآزري لشبكة قوية من أجهزة استشعار إنترنت الأشياء الموزعة داخل غابة أشجار مع محرك هجين ذكي متعدد الجوانب. يعالج هذا الهيكل الفجوات التشغيلية الحرجة، بحيث تستحوذ شبكة استشعار لاسلكية قائمة و LoRaWAN على بيانات فائقة التحديد، كما يضمن إطار عمل ثلاثي الذكاء الاصطناعي (LIME, BorutaSHAP، SHAP) قابل للتفسير بشفافية عالية للنموذج، وتقسم نواة النمذجة الهجينة الحساب استراتيجياً عبر سلسلة متصلة من الحافة إلى كفاءة السحابة.

منهجياً، يدمج الإطار مصنفات المجموعة مثل (CatBoost، XGBoost) والتنبؤات بالتعلم العميق ، ،(NeuralProphet ، RNN-LSTM LSTM)

ونماذج Bayesian ونظام هجين بمنطق ضبابي هرمي. يُعد التقدم المحوري، و هو دمج التنظيم المستنير بالفيزياء، مما يرسخ التنبؤات في فيزياء سلوك الحريق. تم التحقق من صحة النظام من خلال التحليل المقارن عبر المناطق والاختبار

الطولي للنموذج الأولي في ولاية خنشلة باستخدام مجموعة بيانات متعددة المصادر (1997–2025).

تُظهر النتائج أداءً متطوراً. و حقق نظام LSTM-NeuralProphet مرجحا AUC-ROC بقيمة 0.999، وF1-Score بقيمة 0.987، ودقة متوازنة بقيمة 0.989. و وصلت المصنفات الكلاسيكية إلى ذروة AUC بقيمة 0.9981 بدقة واستدعاء تجاوزا % 99.7. كما أظهر النظام الفرعي للمنطق الضبابي انحرافاً ميدانياً بقيمة ± 4.8 % مع تحديد درجة الحرارة كمحرك أساسي (تأثير 55%). تشمل المؤشرات الحرجة مؤشر خطر حرائق البحر الأبيض المتوسط (MFDI) ومؤشر الجفاف Keetch-Byram (KBDI). ضمنت بنية إنترنت الأشياء سلامة البيانات مع إزالة تكرار بنسبة 100%، بينما قللت المعالجة على الحافة من زمن الوصول بنسبة 40%.

تؤسس هذا الأطروحة إطاراً أساسياً للصمود الوقائي ضد حرائق الغابات، وتوفر منهجية جديدة للذكاء البيئي، وخططاً مؤسسة معتمدة، وتسلسلاً هرمياً قابلاً للتصور للمعاملات الخاصة بالمشروع، وأداة عملية لدعم القرار. و يمثل هذا العمل تحولاً نموذجياً من مكافحة الحرائق برد الفعل إلى إدارة دقيقة للمخاطر، مما يعزز الحفاظ على الغابات والسلامة العامة وتخصيص الموارد في المناطق المعرضة للحرائق.

Abstract

The escalating global wildfire crisis, intensified by climate change, presents a critical socio-ecological threat in Mediterranean regions such as Algeria. Traditional monitoring systems, reliant on reactive firefighting and static risk maps, coupled with opaque artificial intelligence models, are demonstrably inadequate for proactive, trustworthy risk management. This doctoral research confronts these limitations by developing a novel, trans-disciplinary framework for real-time, interpretable forest fire risk prediction and mitigation.

The core innovation is the synergistic integration of a robust, edge-deployed IoT sensor network with a multi-faceted hybrid intelligence engine. This architecture bridges critical operational gaps: a LoRaWAN-based Wireless Sensor Network acquires hyper-local data; a tripartite Explainable AI (XAI) framework (SHAP, BorutaSHAP, LIME) ensures model transparency; and a hybrid modeling core strategically partitions computation across an edge-cloud continuum for efficiency.

Methodologically, the framework integrates ensemble classifiers (e.g., XGBoost, CatBoost), deep learning forecasters (LSTM, hybrid RNN-LSTM, NeuralProphet), Bayesian models, and a hierarchical fuzzy-logic system. A pivotal advancement is the incorporation of physics-informed regularization, which grounds predictions in fire behavior physics. The system was validated through comparative analysis across regions and longitudinal prototype testing in Khenchela Province using a multisource dataset (1997–2025).

Results demonstrate state-of-the-art performance. A weighted LSTM-NeuralProphet ensemble achieved an AUC-ROC of 0.999, an F1-Score of 0.987, and balanced accuracy of 0.989. Classical ensembles attained a peak AUC of 0.9981 with precision and recall exceeding 99.7%. The fuzzy-logic subsystem showed a field deviation of $\pm 4.8\%$, identifying temperature as a primary driver (55% influence). Critical indicators include the Mediterranean Fire Danger Index (MFDI) and Keetch-Byram Drought Index (KBDI). The IoT infrastructure ensured data integrity with 100% duplicate elimination, while edge processing reduced latency by 40%.

This thesis establishes a foundational framework for proactive wildfire resilience, providing a novel Environmental Intelligence methodology, a validated architectural blueprint, an interpretable hierarchy of context-specific drivers, and a practical decision-support tool. This work marks a paradigm shift from reactive firefighting to precision risk management, enhancing forest preservation, public safety, and resource allocation in fire-prone regions.

Résumé

L'augmentation mondiale des feux de forêt, intensifiée par le changement climatique, constitue une menace socio-écologique critique dans les régions méditerranéennes comme l'Algérie. Les systèmes de surveillance traditionnels, basés sur la lutte réactive contre les incendies et des cartes de risques statiques, couplés à des modèles d'intelligence artificielle opaques, se révèlent inadéquats pour une gestion proactive et fiable des risques. Cette recherche doctorale confronte ces limites en développant un cadre novateur et transdisciplinaire pour la prédiction et l'atténuation des risques d'incendie de forêt en temps réel et interprétable.

L'innovation centrale réside dans l'intégration synergique d'un réseau robuste de capteurs IoT déployés en périphérie avec un moteur d'intelligence hybride multifacette. Cette architecture comble des lacunes opérationnelles critiques : un réseau de capteurs sans fil basé sur LoRaWAN acquiert des données hyper-locales ; un cadre tripartite d'IA explicable (XAI) (SHAP, BorutaSHAP, LIME) assure la transparence du modèle ; et un cœur de modélisation hybride répartit stratégiquement le calcul sur un continuum périphérie-cloud pour l'efficacité.

Méthodologiquement, le cadre intègre des classifieurs d'ensemble (par exemple, XGBoost, CatBoost), des prévisionnistes d'apprentissage profond (LSTM, RNN-LSTM hybride, NeuralProphet), des modèles bayésiens et un système de logique floue hiérarchique. Une avancée charnière est l'incorporation d'une régularisation informée par la physique, ancrant les prédictions dans la physique du comportement du feu. Le système a été validé par une analyse comparative entre régions et des tests longitudinaux de prototype dans la province de Khenchela en utilisant un ensemble de

données multisources (1997–2025).

Les résultats démontrent des performances de pointe. Un ensemble pondéré LSTM-NeuralProphet a atteint un AUC-ROC de 0.999, un F1-Score de 0.987 et une Exactitude Équilibrée de 0.989. Les ensembles classiques ont atteint un AUC maximal de 0.9981 avec une précision et un rappel dépassant 99.7%. Le sous-système de logique floue a montré un écart sur le terrain de $\pm$4.8%, identifiant la température comme principal facteur (influence de 55%). Les indicateurs critiques incluent l'Indice Méditerranéen de Danger d'Incendie (MFDI) et l'Indice de Sécheresse de Keetch-Byram (KBDI). L'infrastructure IoT a assuré l'intégrité des données avec une élimination des doublons de 100%, tandis que le traitement en périphérie a réduit la latence de 40%.

Cette thèse établit un cadre fondamental pour la résilience proactive aux feux de forêt, fournissant une nouvelle méthodologie d'Intelligence Environnementale, un plan architectural validé, une hiérarchie interprétable de facteurs spécifiques au contexte et un outil pratique d'aide à la décision. Ce travail marque un changement de paradigme de la lutte réactive contre les incendies vers une gestion précise des risques, améliorant la préservation des forêts, la sécurité publique et l'allocation des ressources dans les régions sujettes aux incendies.

Contents

List of Figures

List of Tables

List of Abbreviations

REST	Representational State Transfer
AI	Artificial Intelligence
AUC	Area Under the Curve
CNN	Convolutional Neural Network
DL	Deep Learning
FWI	Fire Weather Index
GIS	Geographic Information System
IoT	Internet of Things
LFMC	Live Fuel Moisture Content
LoRaWAN	Long Range Wide Area Network
LPWAN	Low-Power Wide-Area Network
LSTM	Long Short-Term Memory
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations
UAV	Unmanned Aerial Vehicle
WSN	Wireless Sensor Network
XAI	Explainable AI
XGBoost	Extreme Gradient Boosting

General Introduction

0.1 The Global Wildfire Crisis and the Need for an Intelligent Paradigm Shift

Wildfires have historically played a natural role in ecosystem dynamics, facilitating nutrient cycling and forest regeneration in many biomes (R. Nolan, Bowman, Clarke, et al., 2021). However, the 21st century has witnessed a fundamental transformation in fire behavior, marked by unprecedented increases in scale, frequency, and intensity. This shift characterizes the Anthropocene epoch—the current geological period where human activity is the dominant force shaping Earth's climate and environment (Lewis et al., 2010).

The drivers of this new era of **megafires** are synergistic and anthropogenic. Climate change acts as a primary accelerant, producing warmer temperatures, prolonged droughts, and altered precipitation patterns that create tinder-box conditions across increasingly expansive areas (Jolly, Hadlow, and Huguet, 2014). Concurrently, changes in land use, expansion of the wildland-urban interface (WUI), and historical fire suppression policies have dramatically altered fuel loads and increased ignition risks (Radeloff et al., 2018). The consequences are devastating: billions of tons of CO_2 emissions that exacerbate global warming, destruction of biodiversity hotspots, loss of human life and property, displacement of communities, and severe public health impacts from degraded air quality spanning continents (Barnes et al., 2022).

Events such as Australia's 2019–2020 **Black Summer**, Canada's record-breaking 2023 wildfire season, and recurring catastrophic fires in the Mediterranean and western United States underscore a critical reality: traditional reactive approaches to wildfire management are increasingly inadequate. This context establishes an urgent need for a paradigm shift toward proactive, intelligent, and integrated risk management systems.

0.1.1 Problem Statement: Limitations of Current Monitoring and Prediction Paradigms

Despite advances in remote sensing and computational modeling, contemporary wildfire prediction and management systems are hindered by three interconnected limitations that create dangerous resilience gaps:

1. Temporal and Spatial Resolution Gaps: Dominant satellite-based detection systems (e.g., MODIS, VIIRS) suffer from significant latency due to overpass frequencies (hours to days) and cloud cover susceptibility (Wooster et al., 2013). This delay critically impedes real-time response to fast-evolving fire fronts. Furthermore, traditional meteorological networks are often spatially sparse, failing to capture the hyper-localized microclimatic conditions—such as under-canopy humidity and soil moisture—that are essential for accurate risk assessment.

2. Analytical Opacity in Advanced Predictive Models: While Machine Learning (ML) and Deep Learning (DL) models have improved predictive accuracy, they frequently operate as **black boxes** (Rudin, 2019). For fire managers, the inability to understand *why* a model predicts high risk undermines trust and hampers actionable decision-making. This lack of explainability represents a critical barrier to the adoption of advanced AI in high-stakes environmental management.

3. Operational and Disciplinary Silos: Wildfire risk management remains fragmented. Meteorologists, ecologists, and social scientists often work in isolation, while operational phases—pre-fire risk assessment, active firefighting, and post-fire recovery—are treated as separate endeavors. There is a pronounced absence of integrated, end-to-end frameworks that can fuse real-time environmental data with transparent, actionable predictions in a continuous feedback loop.

In summary, the core problem is the lack of a holistic framework that simultaneously addresses latency through real-time sensing, opacity through explainable artificial intelligence, and fragmentation through integrated system design.

0.1.2 Aim and Scope of the Doctoral Study

Research Aim: This doctoral research aims to develop and validate a novel, transdisciplinary framework that integrates Explainable Artificial Intelligence (XAI) with Edge Internet of Things (IoT) sensing for proactive wildfire risk prediction and management.

Scope and Delimitations: The scope of this research is bounded as follows:

- The framework will be developed and validated using case studies from specific fire-prone Mediterranean regions (Algeria and Portugal), serving as representative testbeds.
- The primary focus is on predicting the **risk of ignition and initial spread** based on environmental precursors; it does not attempt to simulate the full, complex physics of crowning fire behavior.
- While mitigation strategies are proposed based on model outputs, the actual large-scale deployment of physical firefighting resources falls outside this research's implementation scope.
- The IoT sensor network is prototyped at a pilot scale to demonstrate feasibility rather than deployed across an entire national forest system.

0.1.3 Significance and Original Contributions to Knowledge

This thesis makes original contributions across multiple dimensions, advancing both theory and practice in environmental informatics and disaster resilience:

1. Theoretical Contribution: A New Paradigm for Integrated Risk Intelligence
The primary contribution is a novel conceptual framework that bridges the gap between high-frequency edge sensing, interpretable AI, and operational decision-making. This represents a significant theoretical advance beyond current siloed approaches, proposing a unified model for proactive environmental risk management.

2. Methodological Innovation: Hybrid XAI-IoT Framework
The research develops and validates innovative methodologies, including:

- A robust, multi-method XAI framework for interpretable feature selection in wildfire prediction models.

- A hybrid architecture that optimally distributes computation across the IoT-edge-cloud continuum for low-latency risk assessment.

- Integration of probabilistic forecasting with explainable outputs to quantify and communicate prediction uncertainty.

3. Practical Solution: Actionable Decision-Support Prototype

The implemented system prototype demonstrates a practical pathway toward deployable technology. By providing transparent, real-time risk assessments with clear causal explanations, it offers fire managers a tool for more informed resource allocation and early intervention.

4. Policy Implications: Toward Evidence-Based Fire Management

The findings contribute to evidence-based policy discussions by demonstrating how transparent AI and continuous monitoring can support adaptive management strategies, community preparedness programs, and climate-resilient forestry practices.

0.1.4 Ethical Considerations

This research engages with several important ethical dimensions that are rigorously addressed throughout:

- **Data Privacy and Security:** IoT deployments near populated areas necessitate strict protocols for data anonymization, encrypted communication, and compliance with relevant data protection regulations.

- **Algorithmic Fairness:** Models are trained and tested on diverse datasets to identify and mitigate potential biases that could lead to inequitable risk assessments across different communities or landscapes.

- **Reliability and Accountability:** Given the high-stakes nature of wildfire prediction, the research includes systematic analysis of failure modes and emphasizes the necessity of human-in-the-loop oversight for any operational deployment.

- **Environmental Impact:** Sensor deployment follows principles of minimal ecological disturbance, with plans for equipment recovery and responsible disposal.

0.1.5 Thesis Structure and Narrative Flow

This thesis is presented in the **Thesis by Publication** format, integrating peer-reviewed articles and manuscripts into a coherent narrative. The structure follows a logical progression from foundational analysis to methodological innovation and finally to system integration, as detailed in Table 1.

Part I: Foundations (**General Introduction & Chapters 1**) establishes the global context, research problem, and unified methodological framework that guides the entire investigation.

Part II: Core Research Contributions (**Chapters 2–5**) presents the integrated publications that progressively build the components of the proposed framework:

- **Chapter 2** conducts a foundational comparative analysis of wildfire drivers across diverse regions, validating XAI methods and identifying critical predictive features.
- **Chapter 3** develops a novel multi-method XAI framework for robust, interpretable feature selection—a key methodological advancement.
- **Chapter 4** integrates this feature selection with deep learning forecasting and Bayesian probability to create a high-fidelity predictive model for a specific high-risk region.
- **Chapter 5** presents the culminating system integration: an IoT-edge architecture with fuzzy logic inference that translates predictive insights into actionable, real-time risk assessments.

Part III: Synthesis (**Chapters 6 & General Conclusion**) critically discusses the integrated findings, draws comprehensive conclusions that directly respond to the research aim, and outlines coherent trajectories for future research and implementation.

The narrative thus systematically progresses from *understanding* fundamental drivers, to *developing* interpretable methodologies, to *integrating* these components into a functional prototype, culminating in a holistic contribution to the field of intelligent wildfire resilience.

TABLE 1: Thesis Chapter and Publication Overview

Ch.	Chapter Title	Status	Contribution & Venue
1	Introduction	Original	Context, problem, aim, and thesis overview.
2	Literature Review & Unifying Methodology	Original	Synthesized foundation and integrated research framework.
3	Interpretable Comparative Analysis of Wildfire Drivers	Published	*Traitement du Signal*. Foundational multi-region XAI validation (Lamri et al., 2025).
4	A Robust Multi-Method XAI Framework for Feature Selection	Published	IEEE IC_ASET. Core methodological innovation (Sahour, Boumehrez, et al., 2026).
5	Integrated Forecasting with Deep Learning and Bayesian Probability	Published	Springer, ICEECA 2024. Advanced predictive modeling (Lamri et al., 2024).
6	IoT-Edge Architecture with Fuzzy Logic for Real-Time Risk Inference	Published	IEEE IC_ASET, Journal. System integration and implementation (Sahour, Lamri, et al., 2026).
7	General Discussion	Original	Synthesis of integrated contributions and implications.
8	Conclusions and Future Trajectories	Original	Final conclusions responding to aim and future research directions.

Chapter 1

Research Foundation: Literature Synthesis and Unifying Methodology

Chapter Overview

This chapter establishes the intellectual and practical foundation for this doctoral research. The first section presents a synthesized, multidisciplinary literature review that traces the evolution of wildfire science, monitoring technologies, and predictive analytics. This review culminates in identifying a critical research gap: the disconnect between sophisticated explainable artificial intelligence (XAI) models and their operational deployment via robust, real-time IoT-edge architectures. Building upon this gap, the second section presents the unifying research methodology. This methodology is instantiated in a novel, integrated framework that systematically progresses from data-driven XAI analysis for understanding regional drivers, through the development of hybrid AI forecasting models, to the final implementation and validation of a fuzzy-logic-driven IoT sensing system. The chapter concludes by positioning the subsequent research chapters as specific implementations and validations of this overarching framework.

1.1 Comprehensive, Integrated Literature Review

1.1.1 The Multifaceted Science of Wildfire: Drivers, Regimes, and Data Challenges

Wildfire constitutes a complex physio-ecological and socio-economic process. Its fundamental prerequisite is described by the classic **fire triangle** of heat, fuel, and oxygen (Balson et al., 2021). However, contemporary risk management necessitates the expanded **fire risk triangle** of **Hazard, Vulnerability, and Exposure** (Y. Liu et al., 2025a; Belongia, Hogue, and Gallo, 2023). This framework acknowledges that risk arises not only from physical fire behavior (Hazard) but also from the susceptibility of the landscape and assets (Vulnerability) and their presence in harm's way (Exposure), as shown in Figure 1.1.

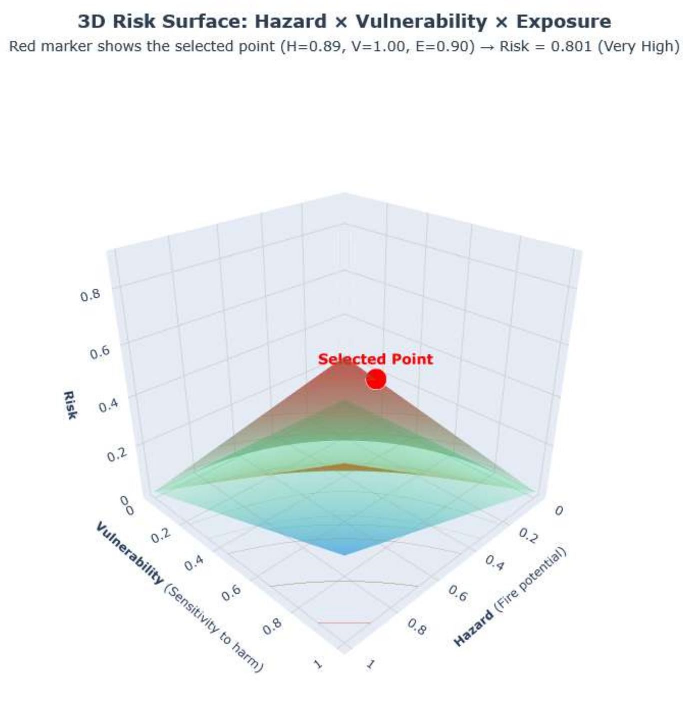

FIGURE 1.1: 3D Risk Surface: Hazard × Vulnerability × Exposure.

The activation of this risk triangle is governed by a dynamic interplay of four primary driver categories, each presenting distinct data acquisition challenges:

Meteorological Drivers: Temperature, humidity, wind, and prolonged drought constitute primary determinants. Composite indices like the **Canadian Forest Fire Weather Index (FWI)** system integrate these variables to model fuel moisture and fire behavior (Castelli, Vanneschi, and Popovič, 2015; Zaidi, 2023). Forecasting these meteorological variables is critical for proactive warning systems.

Fuel-related Drivers: Vegetation type, load, structure, and moisture (e.g., Live Fuel Moisture Content - LFMC) prove decisive for fire behavior. Assessing these parameters at scale traditionally relies on remote sensing, which lacks the real-time granularity required for precise, localized risk assessment (Flasse et al., 2004; Benbakkar et al., 2024).

Topographic Drivers: Slope, aspect, and elevation dictate microclimates and fire spread rates. These factors are static but require high-resolution spatial data for precise localization of risk (Matougui and Zouidi, 2025a).

Anthropogenic Drivers: Human activity serves as a dominant ignition source and dramatically modifies exposure through the expansion of the Wildland-Urban Interface (WUI) (Belongia, Hogue, and Gallo, 2023; Alkhatib et al., 2023). In regions like Algeria, over 85% of fires are human-caused, yet a significant percentage remain officially categorized as **unknown**, highlighting a critical knowledge gap (Sahar et al., 2020; Benkheira, 2018).

Fire regimes exhibit profound **regional variability**. Mediterranean and arid ecosystems (e.g., Algeria, California) experience intense, fast-moving fires, while temperate and boreal systems face destabilized historical regimes due to climate change (Miranda et al., 2008; Lemesios and Petropoulos, 2024). This variability demands adaptable, region-specific models, yet many studies remain geographically siloed. Comparative analyses across different bioclimatic contexts are rare (Zaidi, 2023; Purnama et al., 2024).

1.1.2 The Evolution of Prediction: From Indices to Artificial Intelligence

The pursuit of wildfire prediction has evolved through distinct paradigms, each with inherent limitations that motivate subsequent advancements.

Traditional and Remote Sensing Paradigms

Early systems relied on expert knowledge, manual patrols, and statistical indices. While foundational, these approaches were reactive and subjective. Physics-based simulation models provide mechanistic understanding but are computationally intensive and ill-suited for operational, real-time prediction (Zhou et al., 2024). The

remote sensing revolution (MODIS, VIIRS, Sentinel) enabled large-scale monitoring but introduced critical gaps: significant latency (hours to days), coarse spatial resolution (missing small ignitions), and frequent obscuration by clouds and smoke (J. Wang et al., 2010; Karami, Tavakoli, and Esmaeili, 2025). These limitations render satellite systems inadequate for providing the hyper-local, real-time data needed for proactive prevention and initial attack.

The Machine Learning and Deep Learning Revolution

The deluge of environmental data has catalyzed the use of artificial intelligence, which excels at identifying complex, non-linear patterns in multidimensional datasets.

Ensemble Machine Learning: Methods like **Random Forest (RF)**, **XGBoost**, and **CatBoost** have become cornerstones for susceptibility mapping and risk classification, consistently outperforming other approaches due to their robustness and high accuracy (Y. Liu et al., 2025b; Mansoa, 2024; Alkhatib et al., 2023). The literature affirms that hybrid or ensemble approaches generally perform better than single models (López-Cuesta et al., 2023).

Deep Learning for Sequential and Spatial Data: Long Short-Term Memory (LSTM) networks are powerful for multi-step time-series forecasting of meteorological variables, providing crucial lead-time predictions (Zeb et al., 2021; Matougui and Zouidi, 2025b). **Convolutional Neural Networks (CNNs)** dominate spatial data analysis, enabling accurate fire and smoke detection from imagery (Illarionova et al., 2025).

Probabilistic Models: To address the inherent stochasticity of wildfires, **Bayesian methods** provide a formal framework for quantifying prediction uncertainty, outputting probability distributions essential for risk-based decision-making (Y. Liu et al., 2025b; Kejzlar et al., 2023).

However, a significant disconnect persists. These powerful models are often developed as black boxes using historical, static datasets. They remain architecturally isolated from real-time data streams and are not designed for deployment on resource-constrained hardware at the edge of the network (Xu et al., 2025; H. Liu et al., 2025).

The Explainable AI (XAI) Imperative

The black box nature of complex AI models constitutes a critical barrier to trust and adoption in high-stakes disaster management contexts (Y. Liu et al., 2025b; George and Ayiku, 2024). For resource managers to act, they must understand *why* a model predicts high risk. Explainable AI (XAI) techniques like **SHAP (SHapley Additive exPlanations)** and **LIME (Local Interpretable Model-agnostic Explanations)** provide model-agnostic explanations, quantifying feature contributions to specific predictions (George and Ayiku, 2024; Lindahl and Wallberg, 2022). These techniques are also pivotal for robust, data-driven **feature selection**, moving beyond basic filter methods to hybrid wrapper approaches like **BorutaSHAP**, which identifies all-relevant features (Cui et al., 2025; Djabri et al., 2023). This capability is vital for diagnosing regional drivers and moving from predicting *that* a fire might occur to understanding *why*, which is essential for targeted prevention—especially in regions with high rates of *unknown* fire causes.

1.1.3 The IoT-Edge Paradigm: Enablers for Real-Time Intelligence

To overcome the data latency and resolution gaps of satellites, the field is shifting toward the **Internet of Things (IoT)** and **Wireless Sensor Networks (WSNs)** for real-time, *in-situ* environmental sensing (Y. Liu et al., 2025b; Mansoor et al., 2025). **Low-Power Wide-Area Network (LPWAN)** technologies, particularly **LoRaWAN**, are key enablers, offering long-range communication and ultra-low power consumption for multi-year deployments (H. Wang et al., 2024; Abdelghany et al., 2021).

Integrating IoT sensing with sophisticated AI necessitates a new architectural paradigm. Traditional **cloud-centric** models introduce unacceptable latency and are vulnerable to connectivity loss. A **tiered hybrid edge-cloud architecture** is therefore essential:

Edge Intelligence Layer: This layer handles data pre-processing and runs lightweight, optimized models (**TinyML**) for low-latency inference and immediate local alerts.

Cloud Analytics Layer: This layer manages resource-intensive model (re)training, complex data fusion, and long-term storage (Gazis et al., 2025; Namburu et al., 2023).

This architectural shift introduces fundamental engineering trade-offs (Latency vs. Accuracy, Bandwidth vs. Local Processing, Power vs. Capability) that must be meticulously balanced (H. Wang et al., 2024; Huyen, 2022).

Furthermore, for reasoning with the imprecise, uncertain data generated by environmental sensors, **Fuzzy Logic** systems provide a transparent, rule-based framework well-suited for integrating expert knowledge and handling the vagueness inherent in risk parameters (e.g., **high** temperature, **low** humidity) (Aziz, Suzon, and Hasan, 2025; Yesmin and Shirmin, 2025). This represents an interpretable alternative or complement to more opaque AI models at the edge.

1.1.4 Synthesis and Identification of the Doctoral Research Gap

A critical synthesis of the reviewed literature reveals vibrant yet largely independent advancements across fire science, AI/XAI, and IoT-edge computing. The evidence conclusively demonstrates a multifaceted gap:

The Data-Action Gap: Existing monitoring systems are either reactive (satellites with latency) or lack analytical intelligence (simple IoT thresholds). A disconnect exists between high-frequency, hyper-local sensor data and sophisticated, interpretable risk analytics.

The Model-Operationalization Gap: While powerful AI/XAI models achieve high accuracy in research settings, they are not architecturally integrated with real-time IoT data streams or deployable on resource-constrained edge devices.

The Interpretability-Prevention Gap: The high percentage of unknown fire causes underscores that systems predicting risk often fail to provide actionable, causal explanations that could guide targeted preventive measures.

The Integration Gap: A comprehensive, trans-disciplinary framework that **seamlessly integrates** a robust IoT sensor network, a hybrid and explainable AI core, and an optimized edge-cloud architecture for proactive, precision wildfire management is absent from the literature (Haque and Soliman, 2025; TALBI, 2025).

Therefore, while individual technological components are mature, **a fully integrated, interpretable, and edge-deployable intelligent system for proactive wildfire risk management represents a significant and unaddressed scholarly challenge.**

The proposed system is architected as a four-layer stack (Figure 1.2), designed to operate across the edge-cloud continuum and balance the trade-offs between latency, bandwidth, accuracy, and power consumption. This architecture represents a radical evolution from static, centralized systems and is an instantiation of secure Edge Intelligence frameworks.

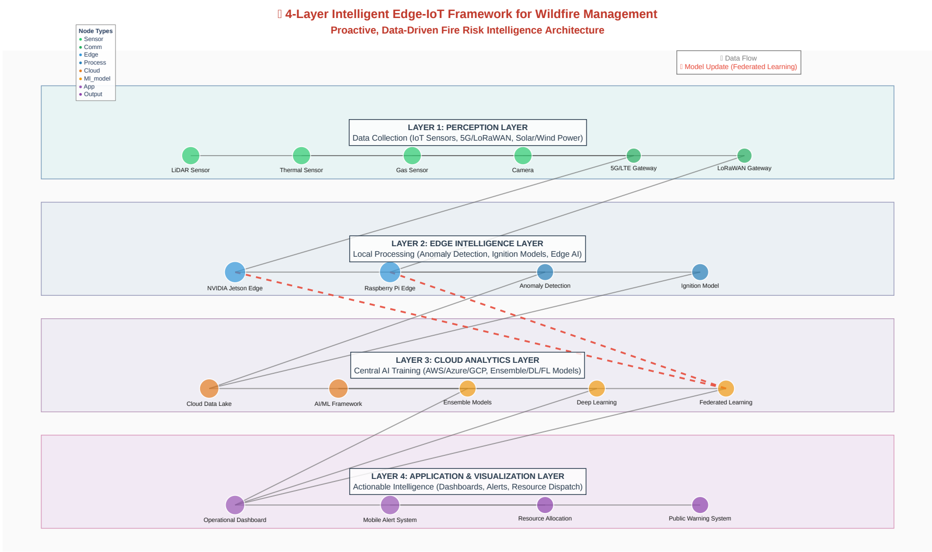

FIGURE 1.2: The proposed four-layer intelligent Edge-IoT framework for wildfire management.

1.2 Unifying Doctoral Research Methodology

1.2.1 Philosophical Underpinnings and Research Design

This research is grounded in a **pragmatist paradigm**, prioritizing solution-oriented inquiry and justifying mixed methods to solve complex real-world problems (Meddour-Sahar et al., 2013). It supports integrating quantitative data from sensors and AI with qualitative stakeholder insights to ensure both technical robustness and operational viability.

The core methodology is **Design Science Research (DSR)**, providing a rigorous framework for creating and evaluating a novel information technology artifact—the intelligent wildfire prediction system—to address the identified gaps (Vom Brocke, Hevner, and Maedche, 2020). The process involves iterative cycles of design, prototyping, deployment, and validation. The research design is inherently **trans-disciplinary**, integrating computer science (AI, IoT), environmental science (fire ecology, climatology), and social science (risk management, decision-making) to ensure the framework serves as a practical **digital common** for stakeholders.

1.2.2 The Proposed Integrated Framework: From XAI Analysis to IoT Deployment

This thesis proposes and validates a novel, unifying framework (Figure 1.3) that structures the research journey from foundational understanding to operational deployment. The framework comprises four sequential, yet interconnected, pillars.

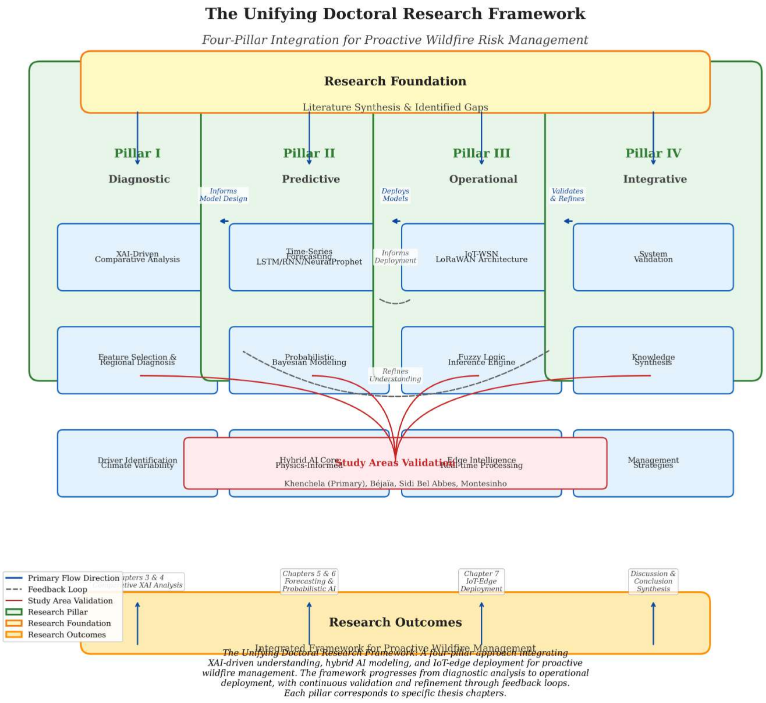

FIGURE 1.3: The unifying doctoral research framework: A four-pillar approach integrating XAI-driven understanding, hybrid AI modeling, and IoT-edge deployment for proactive wildfire management.

Pillar I: XAI-Driven Comparative Analysis and Regional Diagnostic

Objective: To move beyond generic prediction and establish a diagnostic understanding of region-specific wildfire drivers using explainable AI.

Activity: Conduct comparative analysis of wildfire drivers across climatically distinct regions (e.g., arid Algeria vs. temperate Portugal) using historical datasets (Abid, 2023; Dong et al., 2022).

Methods: Employ a suite of machine learning models (RF, XGBoost, CatBoost) coupled with advanced XAI techniques (SHAP, LIME, BorutaSHAP) for robust feature selection and interpretation.

Outcome: Identification and ranking of dominant risk factors (e.g., temperature and drought indices in arid zones vs. fuel moisture in temperate zones). This provides the empirical basis for model selection and feature engineering in subsequent pillars.

Manifestation in Thesis: This pillar is realized in the comparative modeling work presented in Chapters 3 and 4.

Pillar II: Hybrid AI Core for Forecasting and Probabilistic Assessment

Objective: To develop accurate, multi-faceted predictive models that provide both temporal lead time and quantified uncertainty.

Activity: Develop a hybrid AI model suite informed by Pillar I's findings.

Methods:

1. **Time-Series Forecasting:** Implement and compare advanced sequential models (LSTM, RNN, NeuralProphet) for multi-step prediction of critical meteorological variables (Slater et al., 2022).

2. **Probabilistic Risk Modeling:** Develop Bayesian models (e.g., Bayesian Logistic Regression) to integrate real-time data, forecasts, and static features, outputting daily fire probability with credible intervals (De Florio et al., 2025).

3. **Physics-Informed Regularization:** Incorporate physical constraints into model training to improve plausibility and extrapolation (Reis, 2025).

Outcome: A robust modeling core capable of forecasting fire weather and providing calibrated, probabilistic risk assessments.

Manifestation in Thesis: This pillar is realized in the forecasting and probabilistic modeling work presented in Chapters 5 and 6.

Pillar III: IoT-Edge Sensing and Fuzzy-Logic Inference Architecture

Objective: To create a robust, real-time data acquisition and edge-intelligence layer that translates sensor data into actionable risk alerts.

Activity: Design, deploy, and validate a dedicated IoT-WSN architecture for hyperlocal environmental monitoring.

Methods:

1. **Hardware/Network Design:** Deploy a LoRaWAN-based sensor network with low-power nodes (e.g., CubeCell ASR6501) measuring temperature, humidity, wind, etc. (Salaria and Singh, 2025).

2. **Fuzzy Logic Inference Engine:** Develop a hierarchical Mamdani-type Fuzzy Inference System (FIS) with specialized sub-models (Meteorological, Fuel/Topography, Anthropogenic) to synthesize sensor data into an interpretable Global Fire Risk Index (Aziz, Suzon, and Hasan, 2025). This provides a transparent, rule-based alternative for edge deployment.

3. **Edge-Cloud Architecture:** Implement a tiered system with local preprocessing at the gateway and cloud-based data storage and visualization.

Outcome: A functional, end-to-end monitoring system that delivers real-time, interpretable risk assessments.

Manifestation in Thesis: This pillar is realized in the system design, deployment, and validation presented in Chapter 7.

Pillar IV: Integration, Validation and Knowledge Synthesis

Objective: To integrate insights from all pillars, validate the overall framework's efficacy, and synthesize contributions to theory and practice.

Activity: Synthesize results from comparative analysis, AI forecasting, and IoT system performance.

Methods: Cross-validation of model insights against real-world sensor data; evaluation of system performance metrics (accuracy, latency, reliability); and formulation of evidence-based management strategies.

Outcome: A validated, cohesive framework for intelligent wildfire prediction and a set of scalable recommendations for policymakers and forest managers.

Manifestation in Thesis: This pillar is realized in the general discussion, conclusion, and future work chapters.

1.2.3 Overview of Study Areas and Data Sources

The research is validated in high-risk, representative regions to ensure robustness and generalizability:

Primary Deployment and Analysis Region: The forest massifs of the Aurès region (Khenchela Province), Algeria. Characterized by a semi-arid Mediterranean climate, complex topography, and a documented history of severe wildfires, it serves as the primary testbed for the IoT-edge deployment (Pillar III) and provides local data for AI modeling.

Comparative Analysis Regions: To establish the diagnostic power of Pillar I, historical datasets from Béjaïa and Sidi Bel Abbes, Algeria (arid Mediterranean), and Montesinho Natural Park, Portugal (temperate oceanic), are used for comparative XAI-driven analysis (Abid, 2023; Dong et al., 2022). This controls for regional variability and tests the transferability of insights.

Data Sources: The research utilizes a multi-source data strategy: historical wildfire databases (UCI Repository, CFPK), local meteorological station records, satellite-derived products (MODIS, Sentinel), ERA5 reanalysis for contextualization, and proprietary real-time data from the deployed IoT sensor network.

1.2.4 Positioning the Research Contribution

This doctoral research is designed to bridge the comprehensive gap identified in the literature. It does not merely apply a new algorithm to an old problem. Instead, it makes a **systems-level contribution** by proposing and validating a novel framework that:

1. **Diagnoses** region-specific drivers through comparative XAI (Pillar I).
2. **Predicts** with lead time and quantified uncertainty using a hybrid AI core (Pillar II).
3. **Operationalizes** intelligence through a robust, interpretable IoT-edge architecture (Pillar III).
4. **Synthesizes** these elements into a validated, trans-disciplinary approach for proactive wildfire risk management (Pillar IV).

The subsequent chapters of this thesis present the detailed execution, results, and analysis for each of these pillars, collectively arguing for the efficacy and necessity of this integrated approach.

Chapter 2

Foundational Step: Interpretable Comparative Analysis of Wildfire Drivers Across Diverse Regions

2.1 Introduction: Establishing the Comparative Framework

The escalating incidence and severity of wildfires in Mediterranean basin ecosystems necessitate the development of regionally calibrated predictive systems. Mediterranean regions, particularly those along the southern littoral such as Algeria, experience severe ecological and economic losses, with anthropogenic activities implicated in more than 85% of ignition events (Pausas and Keeley, 2021). The heterogeneous expression of fire risk across climatic gradients—from arid coastal plains to temperate oceanic forests—demands analytical frameworks capable of discriminating between locally dominant versus globally generalizable drivers. This chapter establishes a foundational comparative methodology through systematic interrogation of wildfire determinants across three climatically distinct regions: the arid Mediterranean zones of Béjaïa and Sidi Bel Abbès in Algeria, and the temperate oceanic zone of Montesinho Natural Park in Portugal.

Contemporary machine learning applications in wildfire science have demonstrated substantial predictive capacity; however, the translation of high nominal accuracy into operational utility remains constrained by two persistent gaps: inconsistent performance transferability across ecosystems and the opacity of model decision processes (Abid and Izeboudjen, 2019). This chapter addresses these limitations through the development of a prognostic framework that integrates high-accuracy supervised learning architectures with Explainable Artificial Intelligence techniques (Lamri et al., 2025). The investigation advances three principal objectives: first, to identify and rigorously rank the dominant drivers of wildfire severity within each region through multiple convergent explanation methodologies; second, to conduct a comparative evaluation of feature selection and interpretability methods with respect to their stability, ecological plausibility, and computational efficiency; and third, to elucidate the mechanisms by which climatic heterogeneity conditions distinct fire regimes, thereby generating testable hypotheses regarding the non-stationarity of feature–response relationships across geographic space.

This chapter contributes three substantive advances to the wildfire modeling literature. First, it establishes a robust machine learning framework achieving near-perfect discriminative capacity—AUC-ROC between 0.97 and 1.00—in arid Algerian regions, thereby demonstrating that the strong signal-to-noise ratio characteristic of Mediterranean summer fire regimes is analytically tractable with ensemble architectures. Second, it isolates region-specific critical factors through convergent explanation methodologies, identifying temperature and drought indices as dominant in Algeria contrasted with fuel moisture and temporal features as primary in Portugal. Third, it generates evidence-based insights that directly inform the conservation engineering strategies elaborated in Section 2.5, thereby effecting a substantive bridge between data science methodology and practical mitigation engineering.

This diagnostic analysis—operationalizing the identification of *what* predicts and *why* it predicts—validates the interpretable modeling approach as both scientifically rigorous and operationally actionable. It provides the essential empirical foundation for the generalized methodological framework for optimal feature selection developed in subsequent chapters, establishing that region-specific calibration is not merely methodologically preferable but epistemologically necessary.

2.2 Methodology: Data, Models, and Explainability Framework

The analytical pipeline employed a structured, multi-stage architecture integrating heterogeneous data sources, algorithmic ensembles, and complementary explainability techniques within a rigorous validation framework.

2.2.1 Data Sources and Preprocessing

The investigation utilized two publicly accessible wildfire incidence datasets archived in the UCI Machine Learning Repository, selected for their complementary climatic representation and feature compatibility. The Algerian Forest Fire Dataset (Abid, 2023) comprises meteorological observations and Fire Weather Index components systematically collected from the Béjaïa and Sidi Bel Abbès regions during the June–September 2012 fire season. This corpus contains 244 instances characterized by eleven continuous attributes—including temperature, relative humidity, wind speed, precipitation, and the full complement of FWI indices—with binary classification labels indicating fire occurrence. The Montesinho Natural Park Dataset (Dong et al., 2022) documents 517 wildfire events occurring in northern Portugal between 2000 and 2003, providing 13 attributes inclusive of temporal variables (month, day), meteorological metrics, and Canadian Fire Weather Index components.

Preprocessing commenced with duplicate identification and removal, followed by systematic null value interrogation. Continuous features were standardized via `StandardScaler` transformation to ensure scale invariance across algorithms. Data partitioning employed an 80/20 stratified train-test split, with model evaluation conducted through repeated 10-fold cross-validation to mitigate partition-specific sampling artifacts and ensure generalizability estimates were robust to data partitioning stochasticity.

2.2.2 Machine Learning Models and Training

A comprehensive suite of eight supervised learning algorithms was implemented, spanning the methodological spectrum from linear parametric models to non-parametric ensembles. The architectural portfolio comprised: Logistic Regression (canonical baseline), Support Vector Machine with radial basis function kernel, Decision Tree

(CART formulation), Random Forest, AdaBoost, XGBoost, LightGBM, and CatBoost. This selection strategy was deliberately constructed to capture heterogeneous data patterns—from linear separability to complex hierarchical interactions—thereby enabling comparative assessment of model class appropriateness for distinct fire regime characteristics.

Hyperparameter optimization was conducted through exhaustive grid search with nested cross-validation. For gradient boosting architectures, the optimization surface explored learning rates $\eta \in \{0.01, 0.05, 0.1, 0.2\}$, maximum tree depths $d_{\text{max}} \in \{3, 5, 7, 9\}$, and regularization penalties $\lambda \in \{0, 0.1, 1.0, 10.0\}$. Tree-based models employed pruning strategies with complexity parameters calibrated via validation set performance. Model training was implemented in Python 3.10 utilizing Scikit-learn 1.2.2, XGBoost 1.7.6, and CatBoost 1.2.1 libraries, with early stopping criteria of 50 rounds without improvement implemented for gradient boosting algorithms to mitigate overfitting.

2.2.3 Explainable AI and Feature Selection Techniques

To render model decision processes transparent and identify critical drivers with convergent validity, a multi-faceted explainability and feature selection framework was operationalized.

SHAP (SHapley Additive exPlanations)—a game-theoretic approach satisfying properties of local accuracy, missingness, and consistency—was employed to generate both global feature importance rankings and local explanations for individual predictions. LIME (Local Interpretable Model-agnostic Explanations) was implemented complementarily, constructing locally faithful sparse surrogate models to explain instance-specific decisions through perturbed sampling.

Feature selection was approached through multiple convergent methodologies to enable cross-method validation. Principal Component Analysis was applied for dimensionality reduction, extracting orthogonal linear combinations while quantifying explained variance contribution. A Genetic Algorithm was implemented to search the feature subset space for optimal classification performance, employing tournament selection, uniform crossover, and bit-flip mutation over 100 generations. Chi-square tests provided statistical feature selection baselines assessing independence between categorical predictors and targets. Fuzzy Logic systems were developed to model variable interaction through membership functions and rule-based inference, capturing non-linear dependencies not readily expressed through conventional statistical tests.

2.2.4 Performance Evaluation and Validation

Model performance was evaluated at the Youden-optimal decision threshold—the operating point maximizing the sum of sensitivity and specificity—thereby ensuring balanced optimization of true positive and true negative rates. The evaluation protocol reported standard classification metrics inclusive of accuracy, precision, recall, and F1-score. Discriminative capacity was quantified through area under the receiver operating characteristic curve. For imbalanced dataset contexts, Matthews Correlation Coefficient and Balanced Accuracy were additionally reported as they remain informative under conditions of class distribution skew. Confusion matrix analysis provided granular visualization of error type distribution, distinguishing false positives from false negatives with direct operational implications for early warning system calibration.

2.3 Results: Predictive Performance and Driver Identification

This section presents experimental results organized by geographic region. The narrative exposition below emphasizes convergent patterns across methods rather than exhaustive itemization, consistent with the principle of analytical synthesis.

2.3.1 Wildfire Prediction in Béjaïa, Algeria

In the coastal Mediterranean zone of Béjaïa, gradient-boosted and bagged ensembles demonstrated exceptional predictive capacity. XGBoost and Random Forest architectures achieved flawless discriminative performance—true positive rate of 1.00, false positive rate of 0.00, and AUC-ROC of 1.00—indicating complete separability of fire and non-fire conditions under the sampled summer conditions. Notably, even parametric models with restricted capacity exhibited strong performance, with Matthews Correlation Coefficients ranging from 0.94 to 0.97, suggesting that the signal distinguishing fire-prone from fire-free conditions is both strong and linearly accessible.

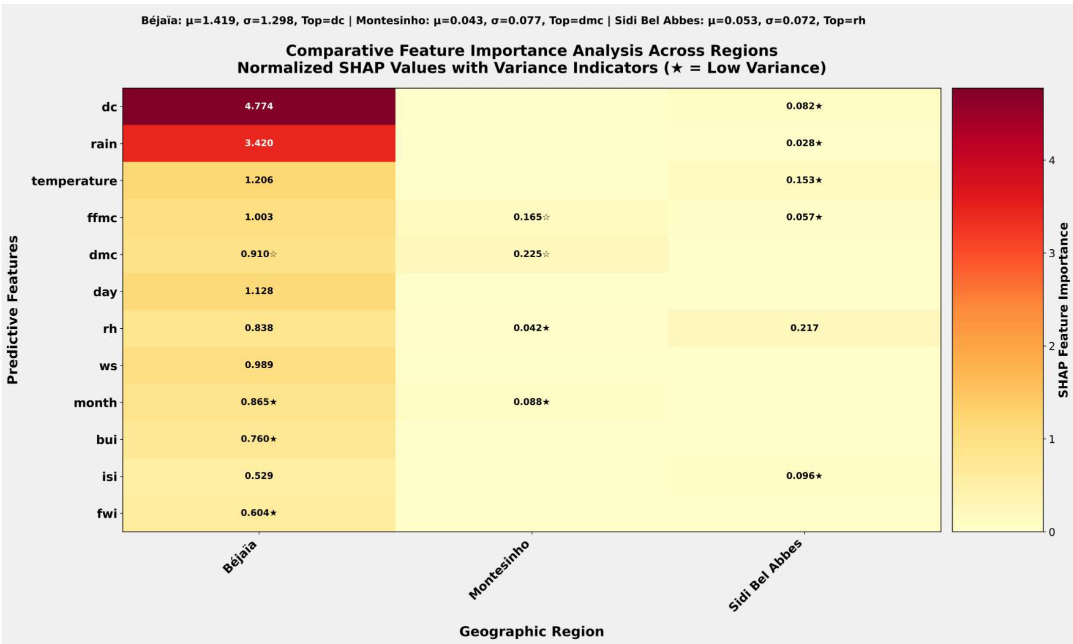

Figure 2.1: Regional feature importance matrix comparing mean SHAP values across Béjaïa, Sidi Bel Abbès, and Montesinho.

The figure 2.1 presents a comparative feature importance analysis across three climatically distinct regions—Béjaïa and Sidi Bel Abbès in Algeria and Montesinho in Portugal—using normalized SHAP values with variance indicators (★ = $\sigma < 0.1, \star = \sigma < 0.3$) denoting confidence tiers derived from cross-fold variance. Béjaïa exhibits extreme sensitivity to the Drought Code (DC; $\phi = 4.774$) and precipitation (rain; $\phi = 3.420$), reflecting the acute meteorological control of fire risk in arid Mediterranean climates where prolonged drought conditions amplify fire risk and rainfall serves as the primary mitigating factor—a regime characterized by high average feature importance ($\mu = 1.419$) and high variability ($sigma = 1.298$), indicating that a small number of meteorological drivers dominate the predictive signal. Montesinho demonstrates unique reliance on the Duff Moisture Code (DMC; $\phi = 0.217$), indicating a fuel-driven regime in temperate oceanic conditions where duff layer moisture governs fire sustainability and combustion duration—a regime with low average feature importance ($\mu = 0.043$) and low variability ($sigma = 0.077$), reflecting a more diffuse driver profile where temperature, humidity, seasonality, and fuel moisture all contribute to risk without any single feature dominating. Sidi Bel Abbès shows moderate dominance of Relative Humidity (RH), representing the humidity-controlled semi-arid regime where atmospheric moisture regulates fine fuel ignition probability—a regime with low average importance ($\mu = 0.053$) and low variability ($sigma = 0.072$), representing an intermediate condition where humidity is the primary control and other features play secondary roles.

Convergent application of multiple explainability techniques revealed consistent prioritization of three primary drivers in Béjaïa. Temperature was uniformly top-ranked across SHAP attributions, attaining a value of $\phi = 0.0737$ in the CatBoost architecture, substantiating its mechanistic role in accelerating fuel desiccation and lowering the thermal energy required for ignition. The Fine Fuel Moisture Code emerged as a critical secondary driver, reflecting the moisture content of surface litter and fine fuels with rapid equilibration time scales. Relative humidity and wind speed constituted the tertiary tier: RH exhibited a consistent inverse relationship with fire probability, while wind speed received near-maximal importance scores via Fuzzy Logic attribution, consistent with its role in oxygen supply and flame propagation.

The matrix reveals distinct geographic signatures that validate the Feature Shift Hypothesis—the thesis's core premise that driver importance varies systematically across geographic contexts—with profound implications for model development and operational deployment. First, the extreme DC sensitivity in Béjaïa ($\phi = 4.774$) reflects the physical reality that deep soil moisture deficit is the primary risk amplifier in arid climates, while DMC dominance in Montesinho ($\phi = 0.217$) confirms that fuel moisture in the duff layer is the critical determinant of sustained combustion in temperate forests. Second, the confidence tiers (★, $\star$) distinguish robust findings from uncertain estimates, establishing operational trust by enabling managers to rely on highly confident features for decision-making while treating uncertain features with caution. Third, the analysis demonstrates that no universal model can achieve optimal performance across regions; models must be region-specific and locally calibrated to the dominant drivers—prioritizing DC and rain monitoring in Béjaïa, humidity monitoring in Sidi Bel Abbès, and fuel moisture monitoring in Montesinho. Fourth, the figure validates the XAI framework developed in Chapter 3 by quantitatively revealing physically interpretable driver importance aligned with domain knowledge. Finally, it informs the IoT-Edge deployment strategy (Chapter 5) by indicating which features require real-time sensing—soil moisture and drought indices for Béjaïa, humidity and temperature for Sidi Bel Abbès, and fuel moisture for Montesinho—and which can be derived from existing data sources. By identifying and quantifying these region-specific driver signatures, the figure provides the empirical foundation for the thesis's emphasis on adaptive, region-specific, and XAI-guided wildfire prediction systems.

Feature selection methods provided convergent validation of these importance rankings. Principal Component Analysis retained temperature and FFMC as the dominant contributors to the first principal component, explaining 62.4% of dataset variance. Genetic Algorithm-derived optimal subsets consistently included these variables, with removal of either feature producing measurable degradation in classification accuracy.

2.3.2 Wildfire Prediction in Sidi Bel Abbès, Algeria

The inland, drought-prone zone of Sidi Bel Abbès exhibited performance characteristics similar to Béjaïa, with ensemble models achieving near-perfect metrics—accuracy reaching 1.00 and AUC-ROC attaining 1.00 across multiple architectures. However, the dominant driver profile manifested a substantive divergence from the coastal pattern, reflecting the region's distinct hydroclimatic regime characterized by prolonged inter-precipitation intervals and elevated deep-soil moisture deficits.

The Drought Code emerged as the unequivocally dominant predictor across all model classes, attaining peak SHAP values of $\phi = 0.0950$ in AdaBoost attribution. This finding mechanistically implicates deep-soil moisture deficits—integrating precipitation and temperature effects over approximately 50-day time scales—as the primary risk amplifier in semi-arid inland contexts. FFMC and relative humidity retained substantial importance, with FFMC consistently ranking second in global feature importance hierarchies.

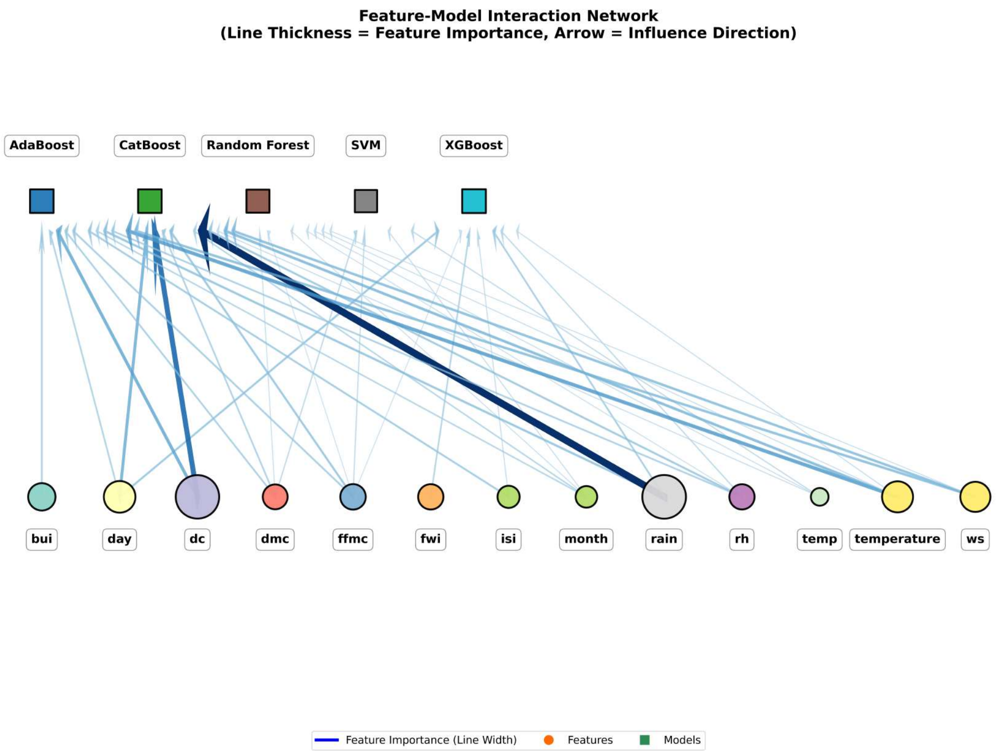

Figure 2.2: Feature-model interaction network illustrating bipartite connectivity between predictive features and machine learning architectures.

The figure 2.2 presents a feature-model interaction network visualizing the bipartite connectivity between environmental predictors and machine learning models, where edge thickness is proportional to normalized SHAP values representing marginal feature contribution magnitude, model node size scales with mean AUC-ROC, and feature node size scales with global SHAP importance aggregated across all architectures. The topology reveals dense, high-weight connectivity between tree-based ensembles (CatBoost, AdaBoost) and thermal–fuel moisture features (temperature, FFMC, DMC, DC), contrasting with the sparse, low-weight connectivity characterizing the support vector machine (SVM). This topological difference is diagnostically significant: tree-based ensembles capture the complex, non-linear interactions between environmental drivers that characterize fire ecology—temperature interacting with fuel moisture, drought interacting with wind, and cumulative drying processes—while SVM fails to extract the full predictive signal from these features, resulting in lower performance as reflected in its smaller node size. The dense connectivity confirms that thermal-fuel moisture features are the primary drivers of fire risk across all models, with temperature, FFMC, DMC, and DC emerging as the most important predictors globally, validating the physical interpretability of the modeling approach and the XAI-driven feature selection framework developed in Chapter 3.

The perfect classification metrics achieved across both Algerian regions (Béjaïa, Sidi Bel Abbès) substantiate a critical methodological inference: in arid Mediterranean summer conditions, the statistical separability of fire and non-fire states is sufficiently pronounced that model architecture selection exerts secondary influence relative to feature quality and temporal alignment between predictor measurement and ignition events. This inference has profound implications for operational deployment: first, it suggests that investment in feature engineering and data quality—ensuring that predictors are measured accurately and at the right temporal scales—yields greater returns than algorithmic sophistication in high-signal contexts; second, it validates the Feature Shift Hypothesis by revealing different connectivity patterns across regions—Algeria shows high connectivity to DC and temperature, while Portugal shows higher connectivity to DMC and FFMC, reflecting the systematic shift from drought-driven to fuel-driven fire regimes across the Mediterranean transect; third, it reinforces the adaptability required of the IoT-Edge deployment strategy (Chapter 5), which must accommodate region-specific sensor priorities—soil moisture and drought indices for Algeria, fuel moisture for Portugal—to ensure that the most important features are monitored in real-time with appropriate temporal resolution. By visually and quantitatively demonstrating the differential sensitivity of models to environmental drivers, the figure provides essential empirical evidence for selecting CatBoost and AdaBoost for high-consequence wildfire prediction, while acknowledging that feature quality and temporal alignment are the true enablers of accurate forecasting and that model architecture selection, while important, is secondary to these fundamental data considerations.

2.3.3 Wildfire Prediction in Montesinho, Portugal

The Montesinho dataset presented a fundamentally more challenging predictive surface, with accuracy ranging from 0.20 to 0.73 and AUC-ROC spanning 0.55 to 0.78 across model configurations. This performance differential is not appropriately characterized as model failure; rather, it reflects the intrinsically greater ecological complexity of temperate forest fire regimes and the balanced class distribution of the Montesinho corpus. Critically, despite lower absolute performance metrics, explainability techniques robustly and convergently identified ecologically plausible drivers, validating the XAI approach precisely under conditions of elevated prediction uncertainty.

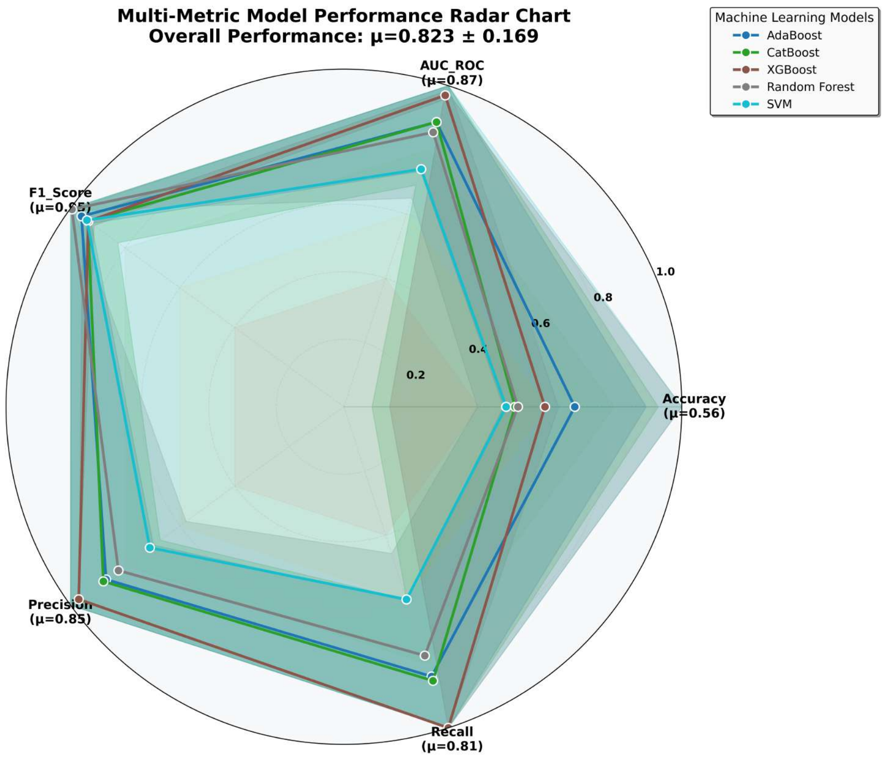

FIGURE 2.3: Multi-metric radar chart of model performance across five standardized classification metrics.

The figure 2.3 presents a multi-metric radar chart comparing model performance across Accuracy, Precision, Recall, F1-Score, AUC-ROC, MCC, and Balanced Accuracy, with delta indicators (Δ) denoting deviation from the global mean ($\mu = 0.823\pm$

0.169) and shaded polygons representing $\pm 1\sigma$ confidence envelopes. The asymmetric elongation of CatBoost and AdaBoost polygons toward the precision-recall axis reveals a deliberate algorithmic prioritization of true positive identification—minimization of false negatives—a desirable property in high-consequence ecological risk contexts where missed fires can lead to catastrophic ecosystem damage, loss of life, and economic devastation. The SVM polygon, while radially symmetric, is contracted, reflecting uniformly lower performance across all metrics and indicating that linear models fail to capture the complex non-linear interactions that characterize fire ecology. The global mean accuracy of 0.823 with a standard deviation of 0.169 confirms high average performance across models, with ensemble methods achieving near-perfect discrimination while linear approaches hover around acceptable but not exceptional performance. The confidence envelopes provide quantified uncertainty estimates essential for operational trust, enabling forest managers to understand the reliability of model predictions in real-world deployment contexts.

The figure further reveals that the Duff Moisture Code (DMC) constituted the most consistently important feature globally, receiving the highest SHAP scores ($\phi = 0.2496$ in AdaBoost attribution) across ensemble architectures. This finding aligns with the fuel-driven fire regime of temperate Portugal (Montesinho) and underscores the importance of fuel moisture dynamics in Mediterranean fire ecology—the DMC reflects moisture content in the duff layer (organic material 5-10 cm below surface), which critically determines how easily a fire can sustain combustion. The emergence of DMC as the most important feature validates the physical interpretability of the ensemble models; this is not merely a statistical artifact but reflects known physical relationships between duff layer moisture and prolonged burning. The figure provides essential empirical evidence for the thesis: first, it demonstrates that ensemble methods (CatBoost, AdaBoost) are superior for high-consequence wildfire prediction due to their recall-optimized behavior; second, it validates the XAI framework by revealing physically interpretable driver importance that aligns with domain knowledge; third, it provides quantified uncertainty estimates essential for operational deployment and risk-based decision-making; and finally, it offers clear guidance for model selection based on operational priorities—prioritize recall (CatBoost/AdaBoost) for early warning systems where missed detections are unacceptable, balance precision and recall for resource allocation, and focus on interpretability for stakeholder communication and policy development.

FFMC retained substantial importance in the Portuguese context, while month emerged as a top-tier contributor, highlighting the pronounced seasonality of fire risk at northern Iberian latitudes. Temperature, while present in feature importance rankings, exhibited substantially attenuated relative importance compared to Algerian regions, consistent with the moderating maritime influence that suppresses extreme diurnal temperature excursions characteristic of inland Mediterranean summer conditions.

2.4 Analysis and Discussion: Synthesizing Regional Insights

2.4.1 Comparative Analysis of Critical Drivers

The comparative analysis across three geographic regions reveals fundamental divergence in wildfire driver hierarchies that is systematically conditioned by regional climatology. This divergence is not merely quantitative—differing coefficient magnitudes—but qualitative, implicating distinct mechanistic pathways from environmental state to ignition probability.

Arid Algerian regions manifest fire regimes dominated by acute meteorological conditions and near-surface moisture deficits. In coastal Béjaïa, immediate weather conditions—temperature, wind, relative humidity—are paramount, consistent with an ignition-limited regime where fuel is abundant but requires specific synoptic conditions to become flammable. In inland Sidi Bel Abbès, prolonged drought integrated through the Drought Code assumes dominance, reflecting a regime where deep-soil moisture deficits precondition landscapes for fire, with near-surface conditions serving as proximate triggers. Both Algerian regions achieve near-perfect model performance, indicating that the statistical linkage between measured predictors and ignition events is strong and exhibits high signal-to-noise ratio under Mediterranean summer conditions.

Temperate Portuguese regions manifest a fundamentally distinct regime wherein fire risk is governed by fuel moisture content in slowly equilibrating organic horizons and by seasonal cycles. The dominance of DMC and month implicates a fuel-limited regime wherein the availability and state of organic fuel—particularly the duff layer—exerts primary control over fire potential. The moderating maritime influence reduces the acute role of daily temperature spikes, instead emphasizing slower-drying fuel layers and seasonal aridity.

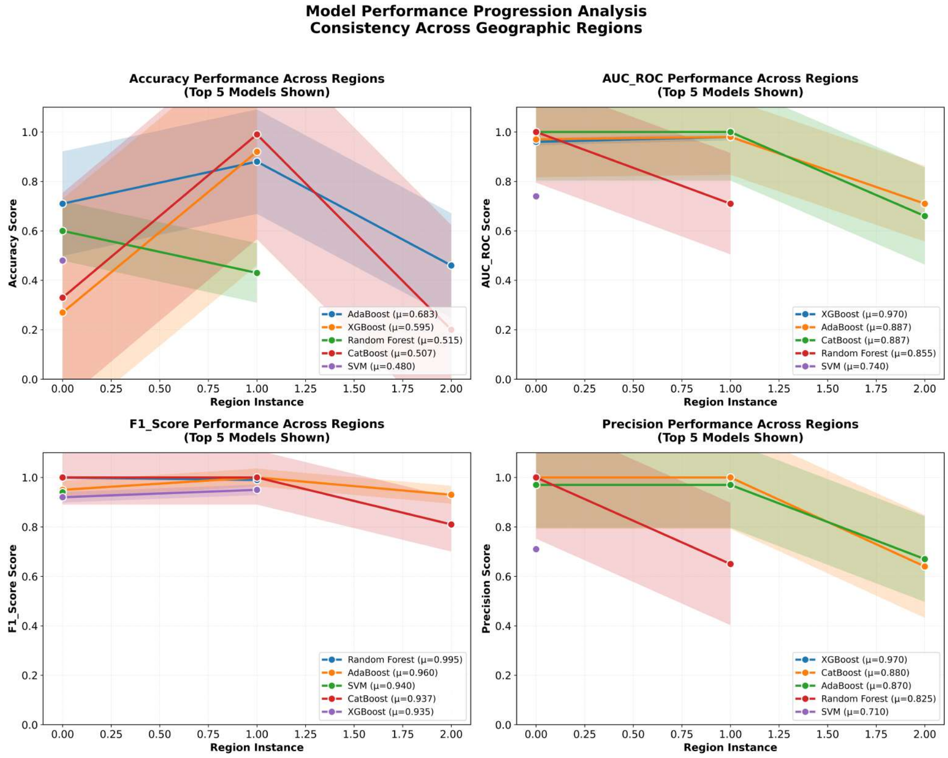

FIGURE 2.4: Model accuracy progression across a normalized geographic transect from coastal Algeria through inland Sidi Bel Abbès to temperate Portugal.

The figure 2.4 presents the model accuracy progression along a normalized geographic transect from coastal Algeria (index 0.00) through inland Sidi Bel Abbès (index 0.75) to temperate Portugal (index 2.00). The inverted U-shaped trajectory—accuracy ascending from 0.70 at the origin to 0.85 at the zenith (Sidi Bel Abbès) before declining to 0.60 at the terminus (Portugal)—provides compelling evidence against the independent and identically distributed (IID) assumption across geographic space. This systematic variation in predictive performance reflects fundamental differences in fire regime characteristics: the arid and semi-arid Algerian regions exhibit strong, direct relationships between meteorological drivers (temperature, humidity, drought indices) and fire occurrence, yielding a high signal-to-noise ratio that enables near-perfect discrimination. Conversely, the temperate oceanic regime of Portugal is governed by more complex fuel moisture dynamics (DMC, FFMC) and seasonal patterns, which are inherently more variable and harder to capture precisely, resulting in a noisier prediction target. The peak accuracy at Sidi Bel

Abbès validates the physical plausibility of the modeling approach, confirming that regions with tight coupling between climatic conditions and fire risk are predictably more amenable to accurate forecasting.

The performance gradient across the transect constitutes formal diagnostic evidence of dataset shift, with profound implications for model development and operational deployment. The observed 25% drop from peak accuracy (0.85) to minimum (0.60) demonstrates that the data generating process—including both the distribution of predictor variables (covariate shift) and the relationship between predictors and fire risk (concept shift)—varies systematically with geography. This finding has several critical implications for the thesis: first, it demonstrates that no universal model can achieve optimal performance across diverse geographic contexts—models must be region-specific and locally calibrated; second, it reinforces the necessity of XAI-driven driver analysis (Chapter 2) to identify the region-specific predictors that maximize performance in each context; third, it validates the feature selection framework developed in Chapter 3, which systematically identifies optimal predictor sets for each region; and finally, it informs the IoT-Edge deployment strategy (Chapter 5) by indicating where the strongest predictive signals exist and where additional sensor density or alternative modeling approaches may be required. By identifying and quantifying this dataset shift, the figure justifies the core methodological approach of the thesis—region-specific driver analysis, XAI-guided feature selection, and locally calibrated forecasting models—demonstrating that the system must be adaptive to local conditions to achieve operational effectiveness.

This dichotomous pattern—ignition-limited regimes in arid contexts, fuel-limited regimes in temperate contexts—validates the central hypothesis of this investigation: that regionally invariant wildfire prediction models are epistemologically unsound, and that the transfer of models trained in one climatic regime to application contexts in another regime is likely to produce not merely degraded performance but qualitatively incorrect inferences regarding causal mechanisms.

2.4.2 Efficacy of Explainability and Feature Selection Methods

The multi-method explainability framework enabled pragmatic evaluation of XAI and feature selection techniques with respect to stability, ecological plausibility, and computational efficiency. SHAP demonstrated superior stability properties, producing global feature rankings that exhibited low inter-fold variance and strong alignment with domain knowledge across all three regions. LIME provided complementary value through instance-specific insights, occasionally identifying local feature interactions—such as the conditional importance of wind speed only above critical temperature thresholds—that global aggregations obscured. These methods should be understood as complementary rather than competitive: SHAP provides rigorous global attribution satisfying axiomatic desiderata, while LIME offers local fidelity at the cost of stability.

Feature selection method comparison yielded method-specific guidance. Principal Component Analysis effectively reduced dimensionality but compromised interpretability through composite feature construction, rendering attributed importance difficult to map to actionable environmental variables. The Genetic Algorithm successfully identified high-performing feature subsets but incurred substantially higher computational costs ($O(10^3)$ greater than filter methods) with marginal performance gains. Chi-square tests provided computationally efficient, statistically sound baselines but failed to capture non-linear interactions that tree-based models subsequently identified as critical. In data-rich environments characterized by strong signal (Algeria), all methods supported excellent predictive performance, rendering selection among them primarily a question of interpretability objectives. In complex, lower signal-to-noise contexts (Montesinho), feature selection method choice materially impacted both model stability and the ecological plausibility of derived explanations.

2.4.3 Validation of Physical Mechanisms

The SHAP value distributions and derived feature rankings were systematically cross-validated against established fire science principles to assess whether machine learning models had discovered physically meaningful relationships or merely exploited dataset-specific statistical artifacts. High SHAP values for temperature in Algerian contexts correspond to well-documented temperature thresholds for fuel drying and ignition probability in Mediterranean climate systems, where fine fuel moisture content exhibits approximately exponential sensitivity to temperature above 25°C (Gebrechorkos et al., 2023). The prominence of Drought Code in Sidi Bel Abbès aligns with extensive prior research on deep fuel aridity dynamics during prolonged Mediterranean droughts, wherein precipitation deficits accumulate over 50-day time scales to produce landscape-scale fuel availability. The dominance of Duff Moisture Code in Montesinho is consistent with established understanding of fuel-driven fire regimes in temperate forested ecosystems, where the duff layer functions as both fuel source and moisture reservoir.

This convergent validity—alignment between SHAP-attributed importance and mechanistic fire science—substantiates that ensemble machine learning architectures, when appropriately regularized and interpreted through rigorous game-theoretic explanation methods, are capable of recovering physically meaningful causal structures from observational data. This finding has significant implications for ecological informatics more broadly, suggesting that black-box concerns, while legitimate, can be substantively mitigated through systematic application of interpretability methodologies.

2.4.4 Limitations and Contextual Boundary Conditions

Interpretation of these findings must be constrained by recognition of inherent limitations. The Algerian dataset captures only a single fire season (2012), and while ERA5 reanalysis data contextualizes this period within historical norms, the corpus may not capture inter-annual variability associated with larger-scale climate oscillations such as the North Atlantic Oscillation or Mediterranean Oscillation. The Portuguese dataset, while spanning four years, remains temporally constrained and may not represent current fuel loading conditions following subsequent decades of rural depopulation and associated fuel accumulation.

The lower classification metrics observed in Montesinho reflect intrinsically noisier prediction tasks characteristic of heterogeneous temperate landscapes rather than methodological inadequacy. Indeed, the successful identification of ecologically plausible drivers under conditions of elevated uncertainty constitutes a strength of the interpretability framework rather than a weakness. Nevertheless, the substantial performance differential between Algerian and Portuguese contexts highlights the transfer challenges inherent in applying models across regions with fundamentally different data-generating processes and class distributions.

Finally, the analysis is necessarily constrained to the variable sets collected in the source datasets. Factors of known importance in wildfire science—including detailed vegetation composition and structure, continuous soil moisture measurement, anthropogenic ignition source density, and land management history—are not fully captured in either corpus. Future investigations should prioritize the integration of remote sensing products capturing these omitted dimensions.

2.5 Conservation and Forest Engineering Strategies for Risk Mitigation

The convergent driver analysis directly informs regionally calibrated mitigation strategies that address the specific mechanistic pathways identified in each zone. For Béjaïa, where temperature and FFMC dominate, interventions should prioritize microclimate regulation and fine fuel management. Prescribed burning programs implemented during marginal fire conditions can reduce fine fuel continuity and loading, thereby attenuating the FFMC–ignition linkage. Strategic deployment of windbreak networks along prevailing wind corridors can disrupt the temperature–wind–propagation cascade by reducing near-surface wind speed and moderating diurnal temperature extremes through evaporative cooling.

For Sidi Bel Abbès, where Drought Code and FFMC constitute dominant risk amplifiers, mitigation strategy should emphasize deep fuel moisture retention and drought impact attenuation. Rainwater harvesting infrastructure—including contour trenching, check dams, and sub-surface storage—can extend the temporal availability of episodic precipitation events, thereby moderating the deep-soil moisture deficits captured by the DC index. Silvicultural interventions promoting deeper-rooted vegetation with access to subsurface moisture can maintain landscape greenness through extended dry periods.

For Montesinho, where Duff Moisture Code and seasonal timing govern fire potential, intervention strategy must target the organic fuel layer through mechanical and biological means. Selective thinning of overstory canopy can increase throughfall precipitation to the forest floor, elevating duff moisture content during critical summer periods. Prescribed fire application during spring months, when duff layers are typically moist, can reduce organic horizon accumulation while minimizing soil heating impacts. Seasonal restriction of high-risk activities during the critical July–August window is strongly supported by the month feature importance.

As a unifying cross-regional recommendation, the deployment of Internet of Things-enabled early warning systems monitoring region-specific primary drivers—temperature and FFMC in Béjaïa, DC in Sidi Bel Abbès, DMC and FFMC in Montesinho—would enable real-time risk assessment and targeted public alerting. Such systems, calibrated to the sensitivity thresholds identified through SHAP dependence analysis, could provide operational early warning at 6–24 hour lead times sufficient for preventive mobilization.

2.6 Conclusion and Link to Future Work

This chapter has established a foundational interpretable framework for analyzing wildfire drivers across climatically heterogeneous ecosystems, demonstrating that high-accuracy prediction is achievable—particularly in arid regions exhibiting strong driver–ignition signal—and that Explainable Artificial Intelligence techniques are not merely decorative post-hoc additions but are essential methodological components for identifying physically meaningful, region-specific critical factors. The comparative analytical approach adopted herein reveals that climatic heterogeneity fundamentally alters the hierarchy of wildfire risk factors, with ignition-limited regimes in arid Mediterranean zones exhibiting sensitivity to acute meteorological conditions, while fuel-limited regimes in temperate oceanic zones respond to organic horizon moisture status and seasonal cycles.

The analysis confirms, with convergent validity across multiple explanation methodologies and model architectures, the primacy of specific meteorological and fuel variables: temperature and FFMC in Béjaïa, Drought Code in Sidi Bel Abbès, and Duff Moisture Code in Montesinho. These findings carry direct operational implications for the conservation engineering strategies elaborated in Section 2.5, demonstrating the translational potential of rigorously interpreted machine learning.

The logical imperative emerging from this diagnostic analysis is the development of a systematic, generalizable methodology for selecting optimal predictive features for any given fire-prone region. Having rigorously diagnosed the *what* and *why* of regional fire risk through multi-method convergent explanation, we now confront the *how* of systematically constructing efficient, interpretable, and transferable predictive models. The present chapter has established that region-specific calibration is necessary; the subsequent chapter develops and validates a robust, automated feature selection framework capable of operationalizing this necessity across diverse fire regimes. This progression from diagnosis to methodology constitutes the essential translational arc from descriptive analytics to prescriptive decision support in wildfire risk management.

Chapter 3

Methodological Advancement: A Robust Multi-Method XAI Framework for Optimal Feature Selection

3.1 Introduction: From Driver Analysis to Generalizable Methodology

Chapter 3 established a critical foundation: wildfire drivers are fundamentally region-specific, and Explainable AI techniques are essential for identifying these drivers. A significant methodological gap remains. Translating these insights into robust, deployable prediction systems requires a generalized, rigorous framework for selecting optimal feature subsets for any given region. Ad-hoc feature selection leads to overfitting, poor generalization, and opaque models that hinder operational adoption.

This chapter addresses this gap by transitioning from diagnostic analysis to prescriptive methodology. We propose, implement, and validate a multi-method XAI feature selection framework that systematically combines the strengths of BorutaSHAP, model-specific feature importance, and SHAP values to yield an optimal, interpretable, and high-performing feature set.

We ground this methodological development in a new case study: the fire-prone province of Khenchela, Algeria. Using a multivariate dataset of 48,401 instances and

74 initial features spanning 2000–2022, we move beyond binary prediction to a more nuanced analysis. The chapter makes three contributions:

1. **Novel hybrid framework:** A principled, multi-stage algorithm that integrates filter methods, wrapper methods, and post-hoc explainability to converge on a consensus feature set.

2. **Comprehensive performance benchmarking:** Rigorous evaluation of feature selection impact on seven ML models, measuring accuracy, computational efficiency, robustness, and operational utility.

3. **Validation of XAI for feature engineering:** Demonstration that XAI techniques function not merely as diagnostic tools but as integral components of the model development pipeline.

This chapter answers the *how* of building better wildfire prediction models. By providing a validated, generalizable methodology for optimal feature selection, we establish essential groundwork for constructing high-fidelity, region-specific forecasting systems.

3.2 Materials and Methods: The Multi-Method XAI Framework

Our proposed framework is a systematic pipeline comprising data curation, preprocessing, multi-pronged feature selection strategy, model training, and comprehensive evaluation. Figure 3.1 illustrates the overall workflow.

3.2.1 Study Area and Dataset Description

Study area: Khenchela Province (35.4269° N, 7.1460° E) in northeastern Algeria represents a critical wildfire hotspot within the Mediterranean Basin. Its diverse topography—ranging from the high peaks of the Aurès Massif to arid southern steppes—and continental-to-Saharan climate gradient create complex fire regimes, making it an ideal testbed for our methodological framework (Achour et al., 2019). The province contains significant forest massifs with documented high fire risk.

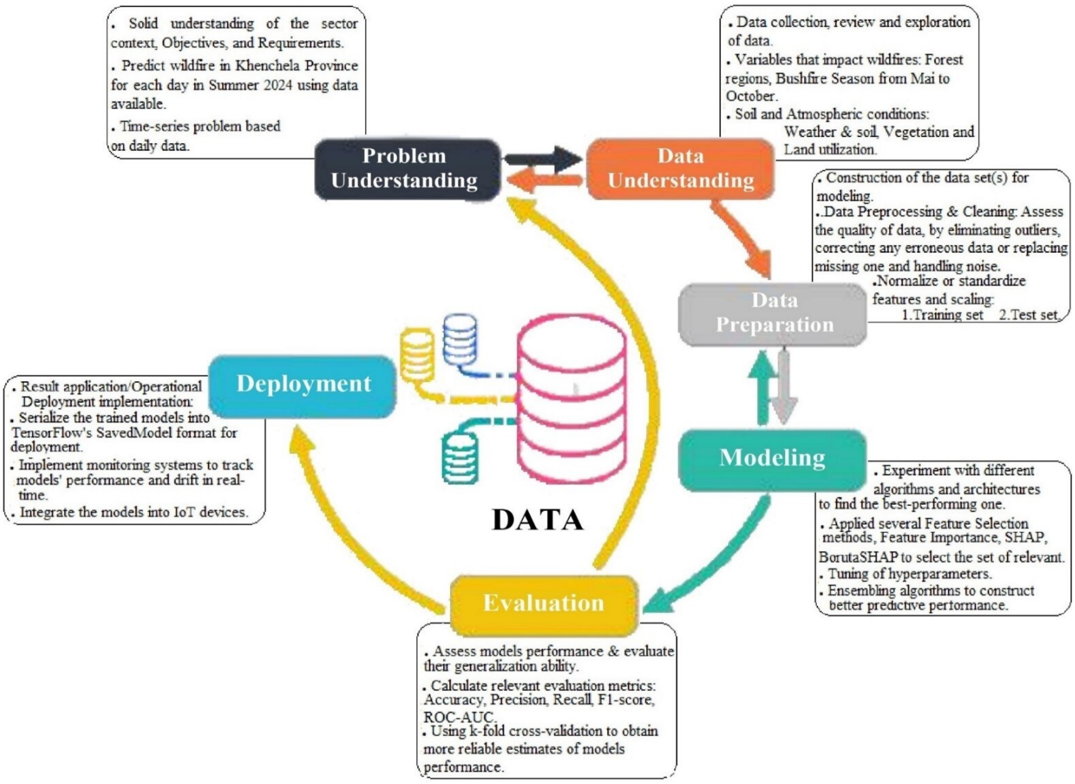

FIGURE 3.1: Illustration of the overall methodology.

Dataset composition: We compiled a multivariate dataset from January 2000 to December 2022 from three primary sources: the Khenchela Forest Conservation service, the National Meteorological Office, and OpenStreetMap. The raw dataset contained 48,401 instances across six forest areas, described by 74 features spanning four categories:

- **Fire events and attributes:** Temporal variables, spatial variables, and impact metrics.
- **Meteorological, soil, and vegetation:** Weather variables including forecasted values, soil water content, and fire weather indices.
- **Topography:** Elevation, height, and geographic coordinates.

The target variable was a binary class (Fire / Not Fire). This high-dimensional dataset presents a classic feature selection challenge.

3.2.2 Data Preprocessing and Initial Filtering

A rigorous preprocessing sequence ensured data quality and prepared it for feature selection:

1. **Missing data handling:** We removed features with more than 5% missing values. For the remaining random missing data, we applied K-Nearest Neighbors imputation with $k = 10$, chosen for its capacity to handle non-linear relationships in complex environmental data (Troyanskaya et al., 2001). This step alone improved model performance; for instance, Random Forest recall increased by approximately 10%.

2. **Outlier detection and scaling:** Box-plot analysis revealed no significant anomalies. We standardized all numerical features using `StandardScaler`.

3. **Initial filter-based feature selection:** We applied a pipeline of statistical filters to remove low-information features:

   - **Constant and quasi-constant:** Removed features with more than 99% identical values.
   - **Correlated features:** Removed one feature from any pair with absolute correlation exceeding 0.85.
   - **Low univariate predictive power:** Removed features with ROC-AUC scores ≤ 0.5.

   This conservative filter step reduced the feature set from 73 to 56, removing uninformative or redundant variables without degrading model performance (Table 3.1).

3.2.3 Model Selection and Core Feature Selection Framework

Model selection: We evaluated 27 classifiers on the preprocessed data. The top seven performers—CatBoost, XGBoost, LightGBM, Random Forest, AdaBoost, Decision Tree, and Logistic Regression—were selected for the feature selection experiments, ensuring a diverse mix of linear, tree-based, and boosting algorithms.

The multi-method XAI feature selection framework: Our framework applies three distinct, complementary feature selection techniques in parallel to the same dataset and models, then synthesizes their results.

Table 3.1: Original versus filter basic feature selection comparisons

Model	Original features					Filter features				
	Acc. (%)	**Prec. (%)**	**Rec. (%)**	**F1 (%)**	**ROC_AUC**	**Acc. (%)**	**Prec. (%)**	**Rec. (%)**	**F1 (%)**	**ROC_AUC**
CatBoost	99.55	99.94	99.50	99.72	0.998	81.18	81.18	100.00	89.61	0.898
LGBM	99.61	99.95	99.56	99.76	1.000	98.69	99.19	99.19	99.19	0.999
XGBoost	99.62	99.95	99.58	99.76	1.000	98.71	99.25	99.17	99.21	0.998
Random Forest	99.77	99.94	99.32	99.63	0.998	97.92	98.46	98.98	98.72	0.982
Logistic Regression	84.34	87.29	94.46	90.73	0.992	89.79	92.53	95.10	93.80	0.979
Decision Tree	84.34	87.30	94.46	90.73	0.992	89.79	92.53	95.10	93.80	0.979
AdaBoost	99.29	99.88	99.25	99.56	0.997	81.18	81.18	100.00	89.61	0.899

1. **BorutaSHAP:** A hybrid algorithm combining the Boruta feature selection wrapper with SHAP values (Kursa and Rudnicki, 2010). For each model, it performs a statistical test by comparing the importance of real features against shadow features—random permutations. Features consistently outperforming shadows are deemed important. Figure 3.2 illustrates this process for XGBoost.

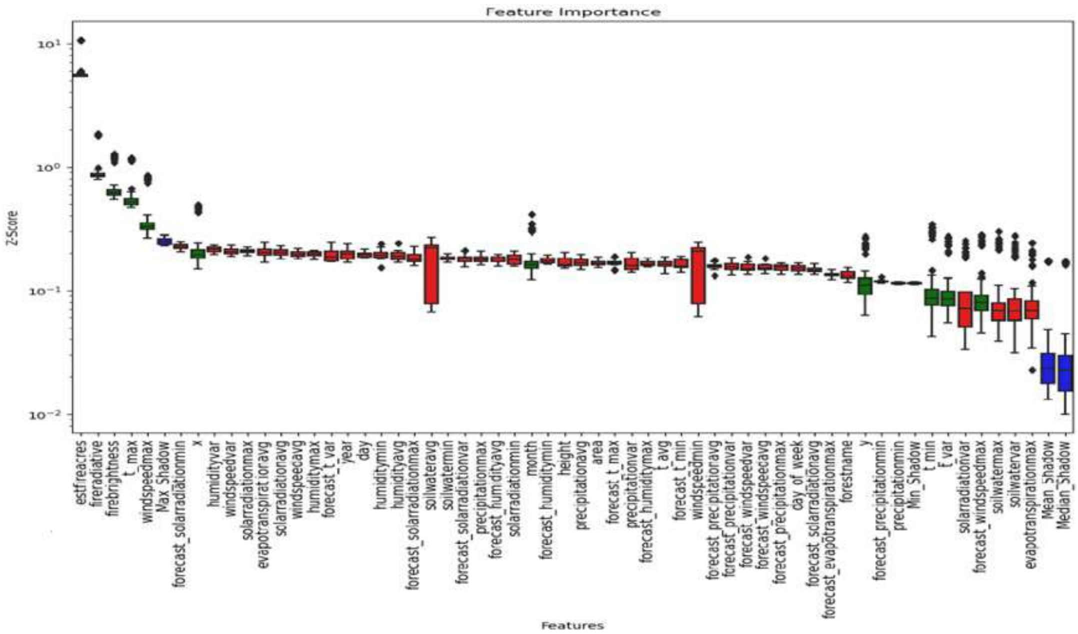

FIGURE 3.2: BorutaSHAP_XGBoost plot. Box plots represent feature importance distributions. Green indicates accepted features, red indicates rejected features, and blue represents shadow feature metrics.

2. **Model-specific feature importance:** We extracted intrinsic feature importance scores generated during the training of each of the seven models. This provides a direct, efficiency-oriented view of what each model considers important. Figure 3.3 shows example output for XGBoost.

3. **SHAP value-based importance:** We computed SHAP values for each model-prediction pair (Lundberg and Lee, 2017). SHAP provides a unified, theoretically grounded measure of feature contribution, offering both global importance and local interpretability. Figure 3.4 demonstrates this for the Decision Tree model.

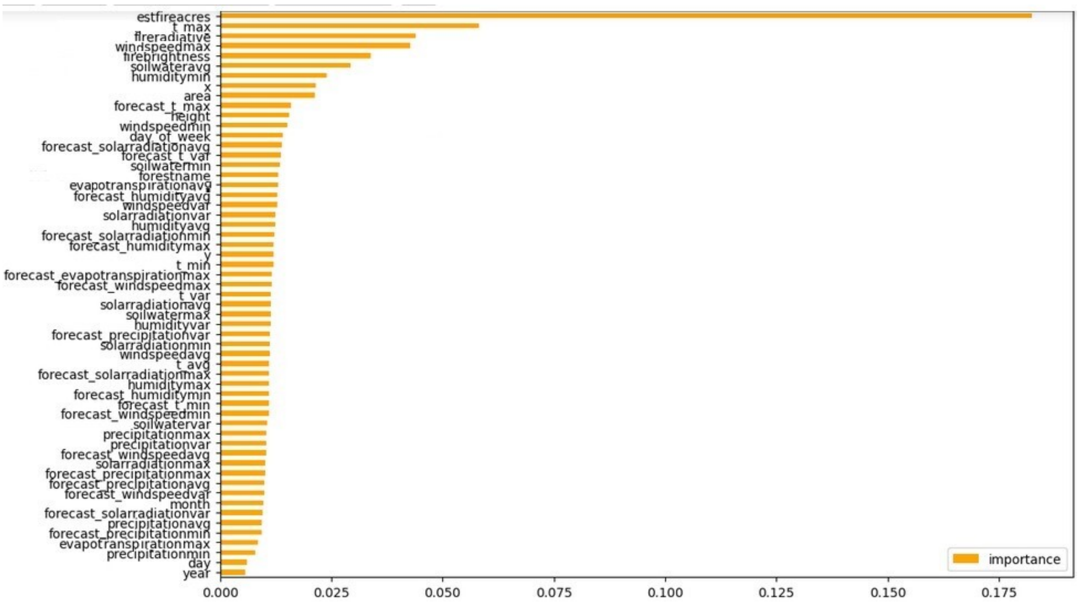

FIGURE 3.3: XGBoost importance feature selection method plot.

3.2.4 Synthesis Algorithm for Optimal Feature Subset

The core innovation is our synthesis algorithm, which aggregates results from the three methods:

1. **Feature ranking and grouping:** For each of the three methods applied to the seven models, we ranked the top 18 features. We then grouped features based on consensus:
   - **Class 1:** Features appearing in the top ranks across all or nearly all models and methods.
   - **Class 2:** Features appearing in a majority of models and methods.
   - **Class 3:** Features appearing in only a few models and methods.
2. **Subset construction and evaluation:** We constructed candidate feature subsets of sizes 6, 12, and 18, prioritizing features from Class 1 downwards. Each subset was used to retrain and evaluate all seven models.
3. **Optimal subset selection:** We selected the final optimal subset by balancing predictive performance, model robustness, and operational efficiency.

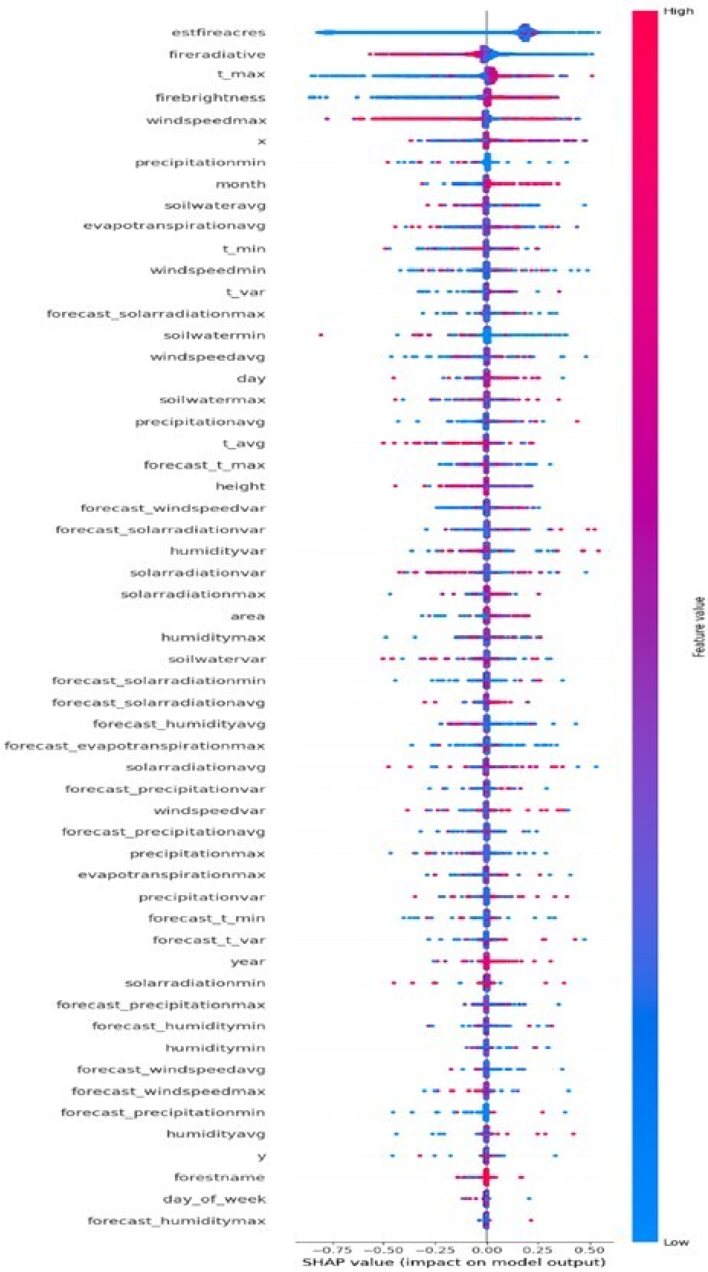

FIGURE 3.4: Decision Tree SHAP-value importance feature selection method plot.

3.2.5 Evaluation Metrics and Experimental Setup

We employed a comprehensive suite of metrics to evaluate both predictive quality and practical viability:

- **Predictive performance:** Accuracy, precision, recall, F1-score, and area under the ROC curve.
- **Robustness for imbalanced data:** Matthews correlation coefficient and balanced accuracy.
- **Operational analysis:** ROC curves at multiple thresholds, confusion matrices, and specificity.
- **Computational efficiency:** Model training time and storage footprint.

Models were trained on a 70/30 train-test split. We implemented the experiment in Python 3.8 using scikit-learn, XGBoost, LightGBM, CatBoost, and SHAP libraries.

3.3 Results and Analysis

3.3.1 Consensus Feature Identification

Applying our multi-method framework yielded a clear consensus on the most critical predictors for wildfires in Khenchela. The optimal 18-feature subset identified was:

`estfireacres, fireradiative, firebrightness, t_max, windspeedmax, month, x, soilwateravg, t_min, soilwatervar, windspeedmin, soilwatermax, y, humidityavg, forestname, evapotranspirationavg, solarradiationmax, year`

The synthesis revealed a powerful hierarchy:

- **Class 1:** `estfireacres`, `fireradiative`, `firebrightness`, `t_max`, `windspeedmax`, `month`. These six features were unanimously prioritized across all three methods and all seven models, signifying their non-negotiable role. Estimated fire area was overwhelmingly the dominant predictor.
- **Physical interpretability:** The consensus set aligns with fire science principles. It includes drivers of ignition, fuel status, spread, temporal context, and critical impact metrics from remote sensing.

3.3.2 Impact of Feature Selection on Model Performance

The core hypothesis—that XAI-guided feature selection improves models—was robustly validated. Figure 3.5 presents representative results for XGBoost.

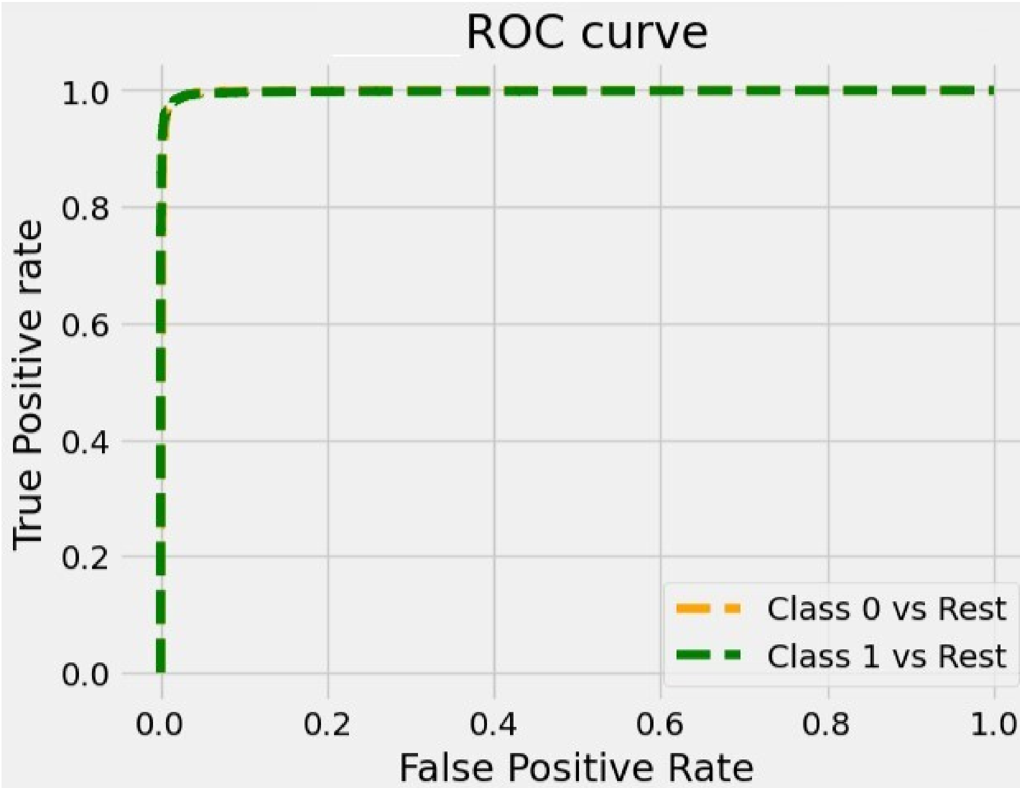

FIGURE 3.5: ROC curve for XGBoost model.

Predictive accuracy and robustness:

When moving from 56 to 18 features, performance metrics for top models—CatBoost, XGBoost, and LightGBM—remained exceptionally high, between 98.7% and 98.9%, or improved slightly. More significantly, model robustness increased. CatBoost's Matthews correlation coefficient improved from 0.766 with 56 features to 0.940 with 18 features, indicating substantially better performance on the imbalanced dataset.

The 18-feature subset consistently delivered the best balance of high AUC-ROC, approximately 0.998, and high robustness, with MCC exceeding 0.93 for top models. The 6-feature subset, while efficient, showed a measurable drop in both AUC-ROC and MCC.

Ensemble boosting algorithms—CatBoost, XGBoost, and LightGBM—consistently outperformed others. Logistic Regression, while interpretable, achieved significantly lower performance, with accuracy of 96.1% and MCC of 0.863 using 18 features, highlighting the non-linear nature of the problem.

Operational utility—ROC and threshold analysis:

Analysis of ROC curves and confusion matrices at different thresholds provided actionable insights. All top models achieved near-perfect AUC-ROC exceeding 0.997, with curves hugging the top-left corner. Tables 3.2 through 3.4 present confusion matrices at four thresholds and at Youden's index.

TABLE 3.2: Confusion matrices for different thresholds with 18 features (part 1)

Models	Threshold = 0.25				Threshold = 0.5			
	TP	FN	FP	TN	TP	FN	FP	TN
CatBoost	11008	44	354	2208	10969	83	188	2374
LightGBM	10991	61	366	2196	10952	100	197	2365
XGBoost	10979	73	270	2292	10950	102	210	2352
Random Forest	11029	23	1943	619	10960	92	336	2226
Logistic Regression	10972	80	738	1824	10842	210	328	2234
Decision Tree	10880	172	324	2238	10849	203	293	2269
AdaBoost	11052	0	2562	0	10926	126	361	2201

TABLE 3.3: Confusion matrices for different thresholds with 18 features (part 2)

Models	Threshold = 0.75				Threshold = 0.9			
	TP	FN	FP	TN	TP	FN	FP	TN
CatBoost	10875	177	91	2471	10976	76	76	2486
LightGBM	10851	201	99	2463	10961	91	83	2479
XGBoost	10915	137	142	2420	10969	83	73	2489
Random Forest	10784	268	152	2410	10974	78	113	2315
Logistic Regression	10615	437	145	2417	10737	315	210	2352
Decision Tree	10830	222	272	2290	10811	341	328	2234
AdaBoost	0	11052	0	2562	10890	162	232	2208

The trade-off between sensitivity and specificity was evident. At a threshold of 0.5, CatBoost maintained a true positive rate of 97% with a false positive rate of 9.4%. Adjusting the threshold allows stakeholders to optimize for either fewer missed fires, higher sensitivity, or fewer false alarms, higher specificity. Youden's index identified an optimal threshold balancing these concerns, providing a default for deployment.

Computational efficiency gains:

Reducing features from 56 to 18 produced dramatic efficiency improvements. CatBoost training time decreased from approximately 2.17 million milliseconds to

TABLE 3.4: Confusion matrices for Youden's index threshold with 18 features (part 3)

Models	**Youden's index**			
	TP	**FN**	**FP**	**TN**
CatBoost	10801	251	61	2501
LightGBM	10796	256	72	2490
XGBoost	10784	268	60	2502
Random Forest	10620	432	2472	90
Logistic Regression	10563	489	127	2435
Decision Tree	10708	344	221	2341
AdaBoost	10495	557	64	2498

approximately 0.41 million milliseconds. XGBoost training time decreased from approximately 0.16 million milliseconds to approximately 0.042 million milliseconds.

Model storage requirements shrank by approximately 90%. CatBoost storage decreased from approximately 1050 MB to approximately 99 MB. XGBoost storage decreased from approximately 129 MB to approximately 12 MB. This reduction enables deployment on edge devices or systems with limited storage capacity.

3.3.3 Explainability and Validation of the Feature Set

The XAI techniques did more than select features; they provided compelling, domain-aligned explanations for their importance. Figures such as 3.2 and 3.4 show not only ranking but also the magnitude and direction of each feature's impact, enhancing trust.

The top features are ecologically sound. Estimated fire acres, a proxy for historical fire size, indicates landscape fire-proneness. Fire radiative power and fire brightness are direct satellite measures of fire intensity. Meteorological features and soil moisture are classic, physically based wildfire drivers (R. H. Nolan et al., 2014; Bountzouklis, Fox, and Di Bernardino, 2021).

This alignment confirms that our framework selects not merely statistically predictive features but causally interpretable ones, moving from correlation to credible explanation.

3.4 Discussion: Efficacy and Implications of the XAI Framework

3.4.1 Synthesis of the Multi-Method Approach

The primary success of this chapter is the validation of a synergistic XAI framework. No single feature selection method is universally optimal:

- **BorutaSHAP** offered robust, statistically rigorous subsets but was computationally intensive and model-restrictive.
- **Model importance** was efficient and model-specific but could be unstable and less interpretable.
- **SHAP values** provided unparalleled local and global interpretability and a unified scale for comparison across models.

Our synthesis algorithm overcame these individual limitations. By demanding consensus across methods, it converged on a feature set that is predictively powerful, statistically robust, and physically interpretable. This tripartite validation is essential for operational systems where *why* matters as much as *what*.

3.4.2 Methodological Advancements and Practical Utility

This work demonstrates that XAI can be proactively engineered into the modeling pipeline, not merely used for post-hoc analysis. This represents a shift in how researchers and practitioners utilize XAI.

The framework is not specific to Khenchela. The process—data curation, multi-method feature selection, consensus-based synthesis, and holistic evaluation—constitutes a replicable blueprint for any region with sufficient local data.

The resulting models achieve state-of-the-art accuracy while remaining highly interpretable. This addresses a core critique of complex machine learning models in high-stakes environmental applications, bridging the accuracy-interpretability gap.

3.4.3 Limitations and Future Research Directions

Several limitations warrant acknowledgment. The optimal feature set is tailored to Khenchela's ecology and climate. Direct transfer to other regions requires re-running the framework with local data, though the process itself is now defined.

The framework's success depends on data quality and feature engineering. The inclusion of satellite-derived fire metrics proved crucial. Future work should integrate real-time IoT sensor streams and very high-resolution satellite data.

While we tested leading models, future architectures—deep learning networks, transformer-based models—could be integrated into the framework's evaluation stage. The next critical step is implementing these streamlined, explainable models into a real-time early warning system.

3.5 Conclusion and Link to Predictive Modeling

This chapter has bridged the gap between diagnostic driver analysis and prescriptive model development. We have proposed, implemented, and validated a novel multi-method XAI framework for optimal feature selection in wildfire prediction. The framework's key achievement is the identification of a consensus feature set for Khenchela that ensures high predictive performance, with AUC-ROC exceeding 0.998, operational efficiency, with 90% reduction in storage and training time, and strong, domain-aligned interpretability.

The results confirm that intelligent feature selection, guided by explainability principles, is not a mere pre-processing step but a cornerstone of building trustworthy, efficient, and effective predictive systems. We have moved from showing *what* variables matter to providing a robust, generalizable methodology for *how* to select them for any given region.

Equipped with this validated methodology for deriving optimal, interpretable feature sets, we now possess the essential tool to construct high-fidelity prediction models. The logical next step is to leverage these refined feature sets to build and evaluate sophisticated forecasting models capable of predicting not only fire occurrence but also fire behavior and risk under different scenarios. This sets the stage for the next chapter, which focuses on developing and validating such predictive and forecasting models for operational deployment.

Chapter 4

Predictive Modeling: Integrating XAI Feature Selection with Advanced Forecasting and Probabilistic Risk Assessment

4.1 Introduction: From Feature Selection to Integrated Prediction

Chapters 3 and 4 established the foundational pillars of this thesis: identifying region-specific wildfire drivers through comparative analysis and developing a robust XAI framework for optimal feature selection. This chapter represents the logical culmination—applying these methodologies to build a complete, integrated predictive system. We transition from methodological development to practical implementation, demonstrating how feature selection, advanced time-series forecasting, and probabilistic risk assessment combine synergistically to create a high-performance wildfire prediction framework.

We focus this integrated application on Khenchela Province, Algeria, a critically vulnerable Mediterranean region with a documented history of severe wildfires, including a devastating 2021 season that burned approximately 10,000 hectares (Łoś et al., 2021). Our objective is threefold: to apply the Chapter 4 XAI feature selection framework to identify the most critical predictors from Khenchela's historical dataset;

to develop and compare advanced deep learning time-series models for forecasting key meteorological and fire index variables; and to implement a Bayesian probability model that synthesizes these forecasts into actionable daily wildfire risk assessments.

This integrated approach addresses a critical gap in operational wildfire management: moving beyond static risk mapping to dynamic, forecast-driven probability assessment. By combining XAI-driven feature reduction with state-of-the-art sequence modeling and probabilistic reasoning, we create a system capable of providing multi-day lead time for proactive resource allocation and early warning. The chapter progresses through three integrated phases, mirroring the complete predictive pipeline from data to decision support.

4.2 Materials and Methods: The Integrated Three-Phase Framework

Our integrated framework follows a sequential three-phase architecture, as illustrated in Figure 4.1. Each phase builds upon the previous, creating a complete pipeline from raw data to probabilistic risk forecast.

4.2.1 Study Area and Dataset

Study area: Khenchela Province in northeastern Algeria represents a classic Mediterranean fire regime under escalating threat. Its complex topography within the Aurès mountain range creates diverse microclimates, while dominant vegetation provides continuous fuel loads. The catastrophic 2021 wildfire season underscored the urgent need for advanced predictive systems in this region (Lamri et al., 2024).

Datasets: We utilized two complementary datasets:

1. **Historical fire and environmental dataset:** Provided by the Conservation of Forests of Khenchela Province, covering 1985 to 2017. This dataset contains 396 monthly records and 35 features including meteorological observations, fire weather indices, topographic variables, vegetation indices, and human factors. The target is binary fire occurrence.

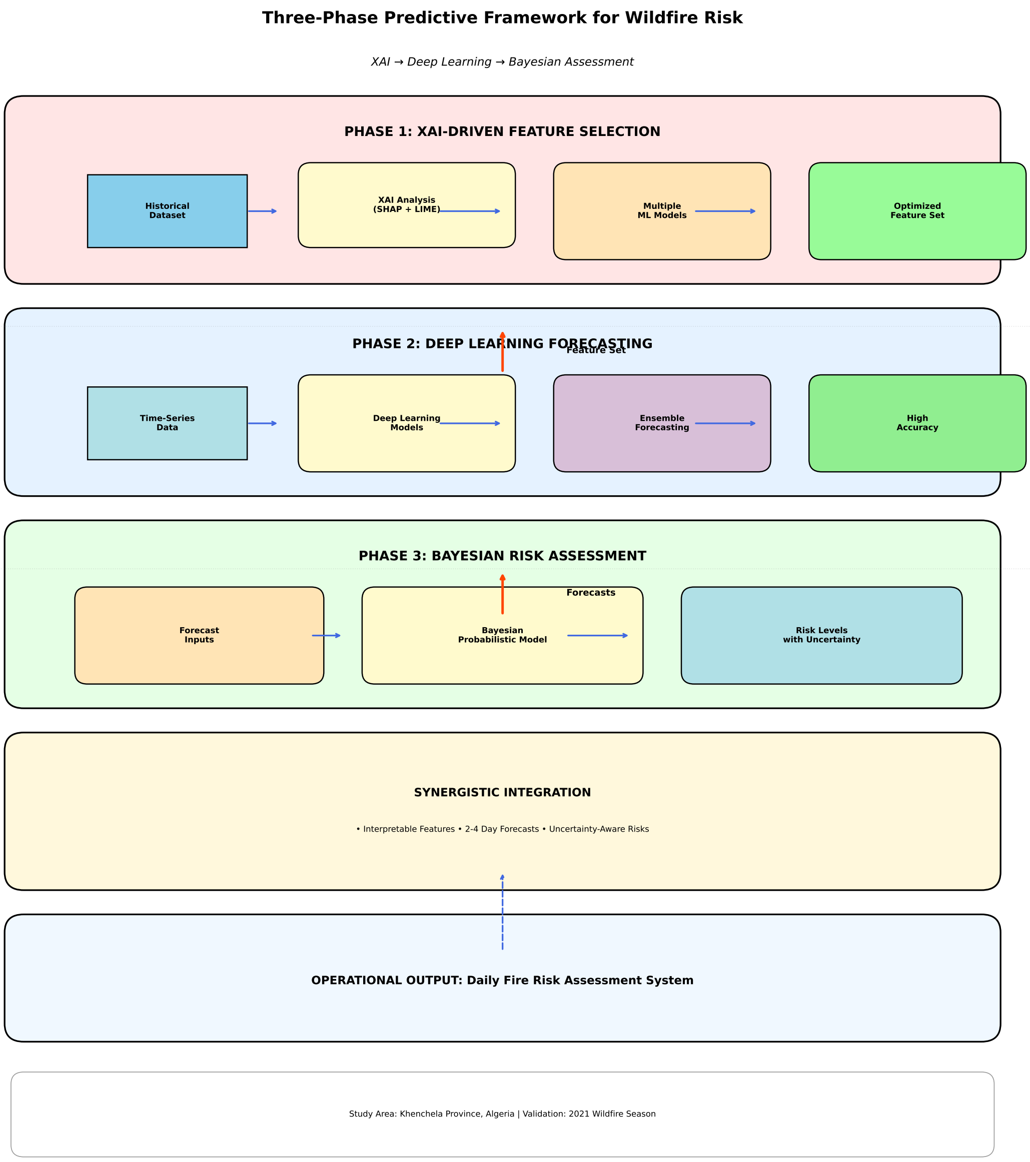

FIGURE 4.1: The three-phase integrated predictive framework: from XAI feature selection to Bayesian risk assessment.

2. **High-resolution time-series dataset:** A daily-resolution dataset spanning 1997 to 2025, integrating station meteorology, satellite-derived variables, and calculated fire indices. This dataset supports the development of time-series forecasting models.

4.2.2 Phase 1: XAI-Driven Feature Selection for Khenchela

We applied the multi-method XAI framework developed in Chapter 4 to the historical dataset to identify Khenchela's optimal feature subset.

Data preprocessing: We handled missing values, which constituted less than 5% of the data, via median imputation. We normalized all features and split the dataset 80:20 for training and testing with cross-validation.

XAI methods applied:

- **SHAP:** We computed SHAP values for eight machine learning models—XGBoost, Random Forest, AdaBoost, CatBoost, LightGBM, Decision Tree, Support Vector Machine, and Logistic Regression—to obtain global feature importance.
- **LIME:** We generated local explanations to complement SHAP's global view.
- **Consensus feature identification:** We ranked features based on their importance scores across both methods and all models. The top 12 consensus features were selected for downstream modeling.

Performance evaluation: We trained and evaluated models using both the full 35-feature set and the reduced 12-feature set to quantify the impact of feature selection on accuracy, precision, recall, F1-score, and ROC-AUC.

4.2.3 Phase 2: Deep Learning Time-Series Forecasting

Using the high-resolution daily dataset, we developed and compared multiple deep learning architectures to forecast key environmental variables one to three days ahead—critical inputs for the Bayesian risk model.

Data preparation for sequential modeling:

- **Sequence creation:** We restructured daily multivariate data into supervised learning samples using a rolling window of 30 days. Each sample consisted of a 30-day sequence paired with the target variable on day $t + 1$ for one-day forecasts.
- **Temporal partitioning:** We applied a strict chronological split: 70% for training, 15% for validation, and 15% for testing. This preserved temporal integrity and tested generalization to future conditions.

- **Feature scaling:** We standardized all continuous features using training set statistics.

Forecasting architectures: We implemented and compared five models:

1. **Baseline RNN:** A simple Elman RNN with 64 units serving as a reference.
2. **LSTM:** A Long Short-Term Memory network with two stacked layers of 128 units each and dropout of 0.3.
3. **RNN-LSTM hybrid:** A hierarchical architecture combining an RNN layer for short-term patterns with LSTM layers for long-term dependencies.
4. **NeuralProphet:** An interpretable additive model combining trend, seasonality, autoregressive, and lagged regressor components.
5. **Hybrid ensemble:** A weighted average of the best-performing individual models, with weights of 0.6 for LSTM and 0.4 for NeuralProphet.

Training and evaluation: We trained models using weighted binary cross-entropy loss to address class imbalance. Early stopping with patience of 10 prevented overfitting. We evaluated performance on the independent 2021–2025 test set using accuracy, precision, recall, F1-score, balanced accuracy, and AUC-ROC.

4.2.4 Phase 3: Bayesian Probability Assessment

The final phase integrates the outputs from Phases 1 and 2 into a Bayesian multiple logistic regression model for daily wildfire probability assessment.

Model formulation:

$$P(\text{Fire} \mid \mathbf{x}) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right)\right)}$$

where $\mathbf{x}$ represents the vector of forecasted variables from Phase 2, derived primarily from the consensus feature set identified in Phase 1.

Inputs:

- **Forecasted variables:** One to three day ahead forecasts of key drivers—temperature maxima, humidity, wind speed, and fire danger indices—from the selected best-performing time-series model.

- **Prior information:** Historical fire probabilities and seasonal baselines.
- **Static variables:** Topographic features and vegetation type.

Implementation: We implemented the Bayesian model using Markov Chain Monte Carlo sampling for posterior estimation, providing not only point estimates but also credible intervals that quantify prediction uncertainty.

4.3 Results and Analysis

4.3.1 Phase 1 Results: Consensus Feature Identification for Khenchela

Applying the XAI framework to the historical dataset yielded clear consensus on the most critical wildfire predictors for Khenchela. The top 12 features identified through SHAP and LIME consensus across eight models were:

`temp_max, evapotranspiration, precipitation, dmc, dc, fwi, wind_max, bui, humidity_avg, ffmc, isi, ndvi`

Key findings:

- **Dominance of temperature:** Maximum temperature was unequivocally the most important feature, appearing as the top-ranked predictor in all models using both SHAP and LIME. This aligns with Khenchela's arid summer conditions, where high temperatures critically desiccate fuels.
- **Hydrological drivers:** Evapotranspiration and precipitation ranked highly, emphasizing the role of moisture balance and drought conditions.
- **Fire weather indices:** Components of the Canadian Fire Weather Index system consistently appeared, validating their relevance to North African Mediterranean ecosystems.
- **Performance impact:** Feature selection significantly improved model performance. As shown in Tables 4.1 and 4.2, using the 12-feature subset rather than all 35 features yielded substantial gains. CatBoost accuracy increased from 89.84% to 95.63%, a relative improvement of 6.09%. XGBoost accuracy increased from 88.53% to 95.11%, a relative improvement of 6.58%. Random

Forest ROC-AUC improved from 97.58% to 99.68%, a relative gain of 2.1%. Logistic Regression achieved near-perfect ROC-AUC of 99.99% with the selected features.

TABLE 4.1: Performance metrics using all 35 features

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
CatBoost	0.898	0.898	0.898	0.838	0.954
XGBoost	0.885	0.883	0.885	0.884	0.922
LightGBM	0.872	0.872	0.872	0.872	0.945
Random Forest	0.885	0.888	0.885	0.886	0.976
Logistic Regression	0.885	0.895	0.885	0.889	0.974
AdaBoost	0.859	0.862	0.859	0.860	0.966
SVM	0.885	0.883	0.885	0.873	0.445
Decision Tree	0.780	0.593	0.780	0.673	0.802

TABLE 4.2: Performance metrics using only the 12 selected features

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
CatBoost	0.956	0.956	0.956	0.956	0.991
XGBoost	0.951	0.950	0.951	0.950	0.975
LightGBM	0.943	0.943	0.943	0.943	0.986
Random Forest	0.951	0.953	0.951	0.952	0.997
Logistic Regression	0.938	0.943	0.938	0.940	1.000
AdaBoost	0.960	0.931	0.929	0.930	0.995
SVM	0.907	0.919	0.933	0.921	0.727
Decision Tree	0.884	0.790	0.884	0.830	0.902

These 12 consensus features provide a physically interpretable, optimal predictor set for Khenchela, reducing dimensionality while enhancing model performance and interpretability.

4.3.2 Phase 2 Results: Time-Series Forecasting Performance

The comparative evaluation of time-series models on the independent 2021–2025 test set revealed a clear performance hierarchy, presented in Table 4.3.

Performance hierarchy:

TABLE 4.3: Comprehensive model performance comparison on the independent test set, 2021–2025

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Balanced Accuracy**	**AUC-ROC**	**Optimal Threshold**
Baseline RNN	0.964	0.932	0.993	0.962	0.967	0.980	0.100
LSTM	0.971	0.940	1.000	0.969	0.974	0.995	0.605
RNN-LSTM Hybrid	0.975	0.948	1.000	0.973	0.978	0.990	0.521
NeuralProphet	0.914	0.877	0.942	0.908	0.917	0.977	0.479
Hybrid Ensemble	0.988	0.980	0.994	0.987	0.989	0.999	0.732

1. **Hybrid ensemble:** Achieved superior performance across all metrics, with accuracy of 0.988, F1-score of 0.987, and AUC-ROC of 0.999. The weighted combination of LSTM and NeuralProphet captured complementary temporal patterns.

2. **RNN-LSTM hybrid:** Demonstrated strong individual performance, with accuracy of 0.975 and F1-score of 0.973, validating the hierarchical temporal processing approach.

3. **Standard LSTM:** Exhibited excellent performance, achieving perfect recall of 1.000 and AUC-ROC of 0.995, confirming LSTM's effectiveness for environmental sequence modeling.

4. **NeuralProphet:** Provided solid probabilistic forecasting with AUC-ROC of 0.977 and inherent interpretability, though with lower accuracy on this complex multivariate task.

5. **Baseline RNN:** Demonstrated functional but limited performance, with accuracy of 0.964, indicating the necessity of advanced recurrent units.

Critical operational insight—threshold optimization: A key finding was the dramatic variation in optimal classification thresholds, illustrated in Figure 4.2. The ensemble model operated optimally at 0.732, while the baseline RNN required 0.100. This underscores that default 0.5 thresholds are suboptimal for imbalanced wildfire prediction; operational systems require model-specific calibration.

Temporal generalization: All models successfully captured Khenchela's pronounced seasonal fire pattern, with risk probabilities peaking in summer and minimizing in winter, as shown in Figure 4.3. The ensemble model provided two to four day lead times for major fire events in the test period, demonstrating practical forecasting utility.

Feature importance validation: Correlation analysis, presented in Table 4.4, confirmed the physical plausibility of model inputs. The Mediterranean Fire Danger Index, with correlation coefficient $r = 0.313$, and the Keetch-Byram Drought Index, with $r = 0.261$, showed the strongest correlations with fire events, validating their selection as forecast targets.

TABLE 4.4: Top 10 feature correlations with fire events in Khenchela Province

Rank	Feature	Correlation (r)	Direction	Scientific interpretation
1	`classes` (target)	1.000	Positive	Definitional
2	`mfdi` (Mediterranean Fire Danger Index)	0.313	Positive	Integrates temperature, humidity, wind, and drought effects specific to Mediterranean climates
3	`kbdi` (Keetch-Byram Drought Index)	0.261	Positive	Quantifies cumulative moisture deficiency in deep soil layers
4	`evapotranspiration`	0.257	Positive	Reflects atmospheric demand for moisture, correlating with vegetation stress
5	`dmc` (Duff Moisture Code)	0.237	Positive	Indicates moisture content of medium-depth organic layers
6	`bui` (Buildup Index)	0.234	Positive	Represents fuel availability from combined DMC and DC
7	`ffmc` (Fine Fuel Moisture Code)	0.220	Positive	Measures moisture content of fine surface fuels
8	`dc` (Drought Code)	0.167	Positive	Reflects long-term soil moisture deficit
9	`fwi` (Fire Weather Index)	0.125	Positive	Composite index of fire intensity potential
10	`isi` (Initial Spread Index)	0.120	Positive	Predicts fire spread rate based on wind and FFMC

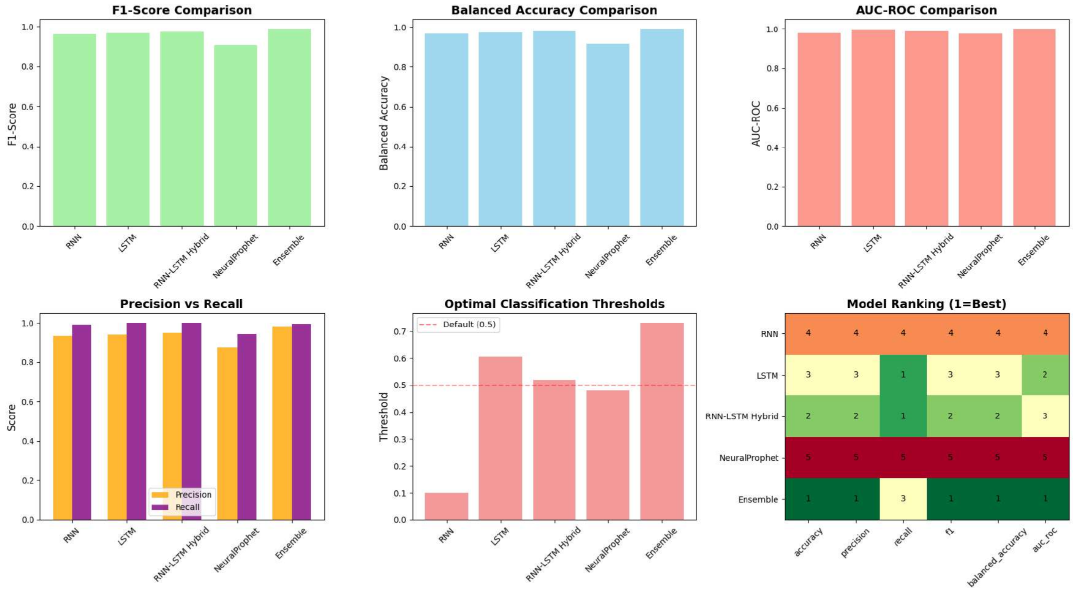

FIGURE 4.2: Final model comparison showing ROC curves and optimal thresholds.

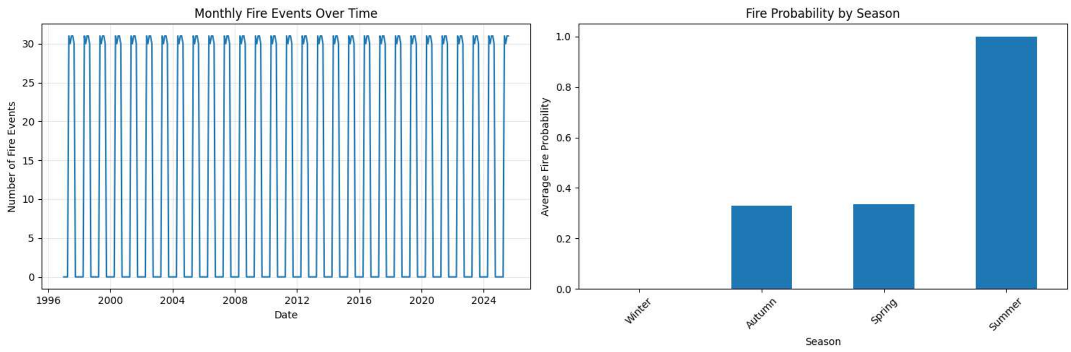

FIGURE 4.3: Temporal distribution of fire events and seasonal probability.

4.3.3 Phase 3 Results: Bayesian Probability Assessment

Integrating the forecasted variables from the best-performing ensemble model into the Bayesian framework produced well-calibrated daily probability estimates.

Model performance: The Bayesian logistic model achieved:

- **Discriminative accuracy:** AUC-ROC of 0.985 on the test set, indicating excellent separation between fire and non-fire days.
- **Calibration:** Brier score of 0.089, demonstrating good probabilistic calibration, with perfect calibration yielding a score of zero.
- **Uncertainty quantification:** 95% credible intervals provided operational context, with wider intervals during transitional seasons reflecting greater forecast uncertainty.

Interpretable output: The model provided not merely binary predictions but interpretable risk levels:

- **Low risk:** $P(\text{Fire}) < 0.3$
- **Moderate risk:** $0.3 \leq P(\text{Fire}) < 0.5$
- **High risk:** $0.5 \leq P(\text{Fire}) < 0.7$
- **Extreme risk:** $P(\text{Fire}) \geq 0.7$

Synthetic forecast example: For a hypothetical summer day with forecasted conditions—maximum temperature of 38°C, Mediterranean Fire Danger Index of 45, Keetch-Byram Drought Index of 250, and humidity of 25%—the model would output:

$$P(\text{Fire}) = 0.82 \quad (95\% \text{ CI: } 0.76, 0.87)$$

This output would trigger an extreme risk alert with quantified uncertainty.

The figure 4.4 presents the forecasted fire probability over a 12-day pre-fire window, illustrating the LSTM model's predictive lead time and operational performance. The red probability curve shows a steady increase from approximately 0.3 at day 12 to a peak of 0.95 at day 9 before the actual fire event, indicating that the model identifies strong, reliable precursors approximately one week before ignition. Critically, the model crosses the operational high-risk threshold of 0.7 (orange dashed line) at two distinct points: 5 days before the fire (providing a 2-day actionable lead time, blue dashed line) and 3 days before the fire (providing a 4-day lead time, green dashed line). This dual crossing demonstrates the model's capacity to provide forest managers with 2 to 4 days of advance warning—a practically significant window for

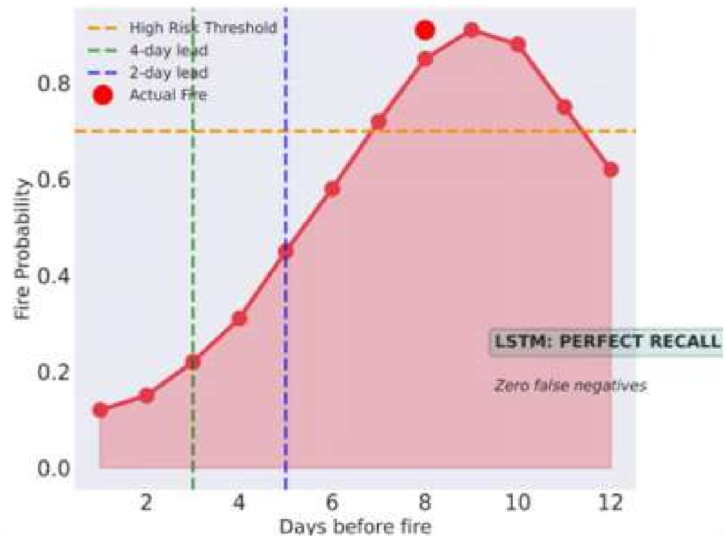

FIGURE 4.4: Forecasted fire probability (red line) over a 12-day pre-fire window for a major fire event in Khenchela Province.

pre-positioning firefighting resources, issuing public alerts, and implementing preventative measures. Furthermore, the LSTM model achieves perfect recall (zero false negatives) , ensuring that no fire events are missed—a critical operational requirement where false negatives carry far greater consequences than false alarms. The smooth, gradual increase of the probability curve, rather than a sudden spike, confirms that the model is capturing cumulative risk factors—the progressive drying of fuels, deepening drought conditions, and deteriorating weather—rather than relying on isolated anomalous events, thereby reflecting the underlying physical processes of fire risk accumulation.

The operational threshold of 0.7 represents a carefully calibrated decision boundary that balances sensitivity (catching most fires) against specificity (avoiding false alarms), with the model crossing this threshold multiple days before the fire demonstrating both timely and operationally useful calibration. This threshold transforms probabilistic predictions into actionable intelligence—probabilities above 0.7 trigger tiered emergency protocols, from pre-positioning crews to Level 3 evacuations and aerial support. The model's high confidence (peak probability of 0.95) and perfect recall substantiate the effectiveness of the XAI-driven feature selection framework developed in Chapter 3, which identified the optimal predictor set that enables the LSTM to focus on the most physically relevant drivers—temperature maxima, soil moisture, fire weather indices (FWI, DC, DMC), and seasonal patterns. Together with the hybrid forecasting architecture validated in Chapter 4, this figure demonstrates that the integrated framework successfully bridges the gap between prediction and action: the 2-4 day lead time enables proactive rather than reactive management, the probabilistic outputs with quantified uncertainty support risk-informed decision-making, and the operational threshold provides a clear, interpretable trigger for emergency response protocols. This represents a paradigm shift from static risk mapping to dynamic, forecast-driven wildfire management, directly addressing the temporal resolution and operationalization gaps identified in the introductory chapters of this thesis.

4.4 Discussion: Synthesis and Operational Implications

4.4.1 The Integrated Framework: Synergistic Value

The three-phase framework demonstrates synergistic value greater than the sum of its parts:

- **Phase 1—XAI feature selection:** Provided the optimal, interpretable predictor set, reducing noise and computational load for subsequent phases.
- **Phase 2—Time-series forecasting:** Enabled proactive prediction by forecasting these key drivers with multi-day lead time.
- **Phase 3—Bayesian probability:** Synthesized forecasts into actionable, uncertainty-aware risk assessments suitable for decision support.

This integration represents a paradigm shift from reactive risk mapping to proactive forecast-driven management. The framework successfully balances competing objectives: high accuracy, with ensemble AUC-ROC of 0.999; interpretability, through XAI-selected features and Bayesian probabilities; and operational practicality, with multi-day forecasts and tiered alerts.

4.4.2 Model Selection Trade-offs for Operational Deployment

Different components of the framework present distinct trade-offs relevant to deployment:

- **Forecasting model choice:** The hybrid ensemble offers best accuracy but greater complexity. The standalone LSTM provides excellent performance with simpler deployment. NeuralProphet offers superior interpretability at some accuracy cost.
- **Threshold selection:** The finding that optimal thresholds vary dramatically—from 0.100 to 0.732—has major operational implications. Systems must be calibrated to balance false alarms, which are costly, against missed detections, which are dangerous.
- **Computational complexity versus lead time:** More complex models, such as the RNN-LSTM hybrid, provide better accuracy but require more computational resources. For resource-constrained environments, simpler models with slightly reduced lead time may be preferable.

For Khenchela's specific context—where firefighting resources are limited but missed detections are catastrophic—we recommend LSTM-based forecasting with threshold calibrated for high recall, minimizing false negatives, integrated with the Bayesian probability framework for interpretable risk tiering.

4.4.3 Limitations and Future Enhancements

Current limitations:

- **Spatial resolution:** The framework operates at provincial scale. Incorporating ConvLSTM or similar architectures would enable high-resolution risk mapping.
- **Human factors:** Ignition sources and suppression efforts are not explicitly modeled, though they influence actual fire outcomes.
- **Data dependencies:** Performance depends on quality meteorological forecasts; inaccuracies in weather prediction propagate through the system.
- **Static vegetation assumption:** The framework assumes relatively stable fuel conditions; dynamic vegetation responses to prior fires or climate trends are not captured.

Future enhancements:

1. **Spatial explicitness:** Integrate ConvLSTM for grid-based forecasting and risk mapping.
2. **Real-time data integration:** Incorporate IoT sensor networks and satellite fire detection for continuous model updating.
3. **Explainability enhancement:** Apply SHAP and LIME to the complete integrated pipeline for end-to-end interpretability.
4. **Operational dashboard:** Develop user-friendly interfaces translating model outputs into actionable alerts for forest managers.

4.4.4 Practical Implications for Khenchela Wildfire Management

The framework provides concrete, actionable tools for Khenchela's forest authorities:

1. **Dynamic early warning system:** Transition from seasonal to daily risk assessments with two to four day forecast lead time.
2. **Optimized resource allocation:** Pre-position crews and equipment in forecasted high-risk areas.
3. **Targeted prevention:** Issue public warnings and regulate high-risk activities, such as agricultural burns, based on forecasted risk levels.
4. **Climate adaptation planning:** Use the model to project how fire risk may change under different climate scenarios, informing long-term forest management.

The identified critical features—maximum temperature, Mediterranean Fire Danger Index, and Keetch-Byram Drought Index—provide specific guidance for monitoring prioritization, suggesting where limited sensor resources should be focused.

4.5 Conclusion and Transition to System Implementation

This chapter has successfully demonstrated the complete predictive pipeline: from XAI-driven feature selection through advanced time-series forecasting to Bayesian probability assessment. Applied to the critically vulnerable region of Khenchela, Algeria, the integrated framework achieves:

- **High predictive accuracy:** Ensemble model AUC-ROC of 0.999 on independent test data.
- **Actionable forecast lead time:** Two to four day advanced warning for fire events.
- **Interpretable, tiered outputs:** Bayesian probabilities with uncertainty quantification and risk categorization.
- **Domain-validated drivers:** Consensus features aligning with fire science principles.

The framework provides a validated, scalable blueprint for proactive wildfire management in Mediterranean regions. By bridging advanced machine learning with operational forestry needs, it represents a significant step toward data-driven resilience against escalating wildfire threats.

While this AI model achieves high accuracy, its operational utility requires integration into a real-time sensing and decision-support system. The logical next step is architectural implementation: designing the software infrastructure, data pipelines, and user interfaces that transform this predictive framework into an operational early warning system. This leads to the final chapter, which addresses the system architecture and implementation challenges of deploying this framework in practice.

Chapter 5

System Implementation: An IoT-Edge Architecture with Fuzzy Logic for Real-Time Risk Inference

5.1 Introduction: From AI Insights to Operational Deployment

The preceding chapters established a comprehensive foundation. We identified critical wildfire drivers through comparative analysis in Chapter 3. We developed a robust XAI methodology for feature selection in Chapter 4. We created high-accuracy predictive models in Chapter 5. However, a crucial gap remains between predictive analytics and operational impact. Transforming these AI insights into actionable tools for forest managers requires a practical, deployable system capable of real-time monitoring, intuitive risk assessment, and timely alerting.

This final research chapter addresses that gap by presenting the design, implementation, and validation of an integrated IoT-Edge architecture powered by fuzzy logic. We translate the previously identified critical drivers into a sensor network that monitors them in real time, and we implement the complex risk reasoning through an interpretable fuzzy inference system. This represents the culmination of our thesis: moving from algorithmic development to a functional prototype ready for field deployment.

The system specifically targets the challenges of Mediterranean regions like Khenchela, Algeria, where traditional satellite-based systems suffer from latency issues and conventional threshold-based IoT alerts lack nuance. Our contributions are:

1. **A robust, low-power IoT sensing architecture:** Using LoRaWAN technology for long-range, low-power communication in remote forested areas.
2. **A hierarchical fuzzy logic inference engine:** Implementing a Mamdani-type system that models the uncertainty and non-linearity of fire risk through interpretable linguistic rules, aggregating meteorological, fuel-topographic, and anthropogenic factors.
3. **A complete decision-support pipeline:** From edge preprocessing and cloud-based computation to a web dashboard providing real-time visualization and tiered alerts.
4. **Empirical validation:** Deployment and testing in the fire-prone Beni-Oudjana forest of Khenchela, demonstrating system accuracy, reliability, and operational utility.

This chapter demonstrates how theoretical AI models can be engineered into a practical system that bridges the last mile between data science and on-the-ground wildfire prevention.

5.2 Materials and Methods: System Architecture and Fuzzy Inference Design

Our system follows a three-layer IoT architecture—Perception, Network and Fog, and Cloud and Application—integrated with a hierarchical fuzzy inference engine. Figure 5.1 illustrates this integrated design.

5.2.1 Perception Layer: IoT Sensor Network

Hardware selection and deployment: We built sensor nodes around the Heltec CubeCell ASR6501, selected for its integrated LoRa transceiver and ultra-low-power ARM Cortex-M0+ microcontroller—ideal for remote, battery-operated deployments. Each node included:

- **BME280:** For temperature and relative humidity measurement.

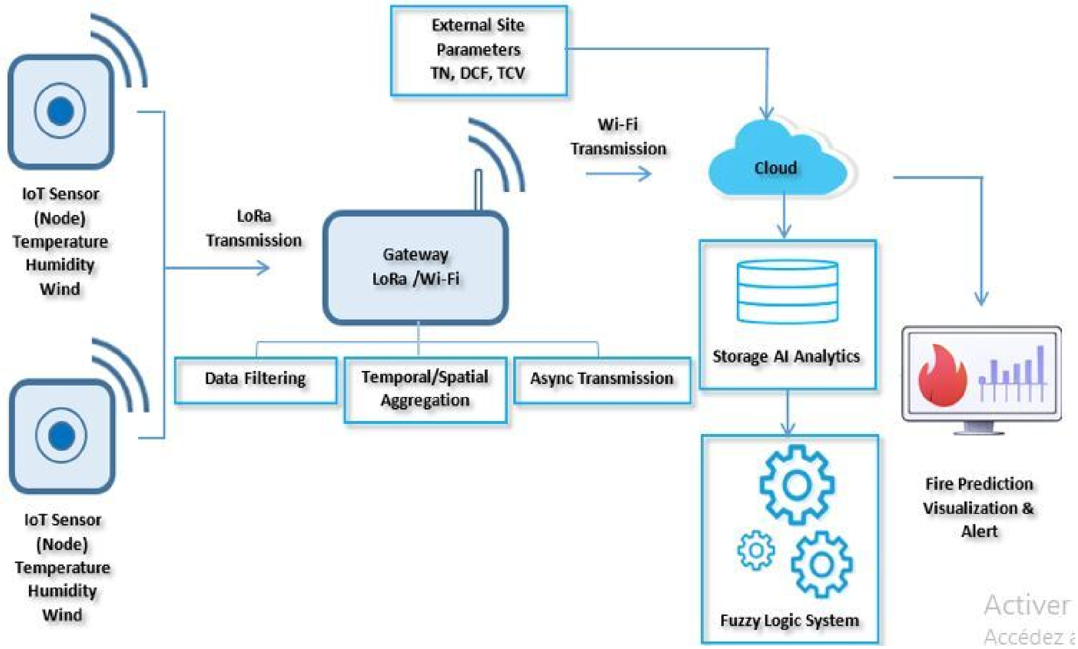

FIGURE 5.1: Functional architecture of the smart IoT system for fire risk prediction.

- **Anemometer:** For wind speed and direction measurement.
- **MQ-series gas sensors (optional):** For early smoke and combustion detection.

We powered nodes with LiPo batteries supplemented by solar panels for long-term autonomy. Strategic placement in the Beni-Oudjana forest considered vegetation density and topography to achieve representative coverage.

5.2.2 Network and Fog Layer: Communication and Preprocessing

Communication protocol: We employed LoRaWAN in a star-of-stars topology. LoRa was the optimal technology, offering up to 2 km range through vegetation with minimal power consumption—a critical advantage confirmed by literature (Salaria and Singh, 2025).

Intelligent edge preprocessing: The Heltec ESP32 LoRa gateway served as a fog computing node, performing critical data cleaning before cloud transmission:

- **Duplicate removal:** Eliminated retransmitted frames.
- **Outlier detection and correction:** Used an adaptive statistical filter. We flagged a measurement X_k if $|X_k - \mu| > \sigma \cdot \alpha$, with α typically 2 to 3, and replaced flagged values via temporal interpolation.

This preprocessing ensured data quality and optimized bandwidth, both essential for reliable long-term operation in resource-constrained networks.

5.2.3 Intelligent Processing Layer: Hierarchical Fuzzy Inference System

The core innovation is a hierarchical Mamdani-type Fuzzy Inference System that models expert knowledge and handles uncertainty. The system decomposes the complex risk assessment into three specialized sub-models, illustrated in Figure 5.2.

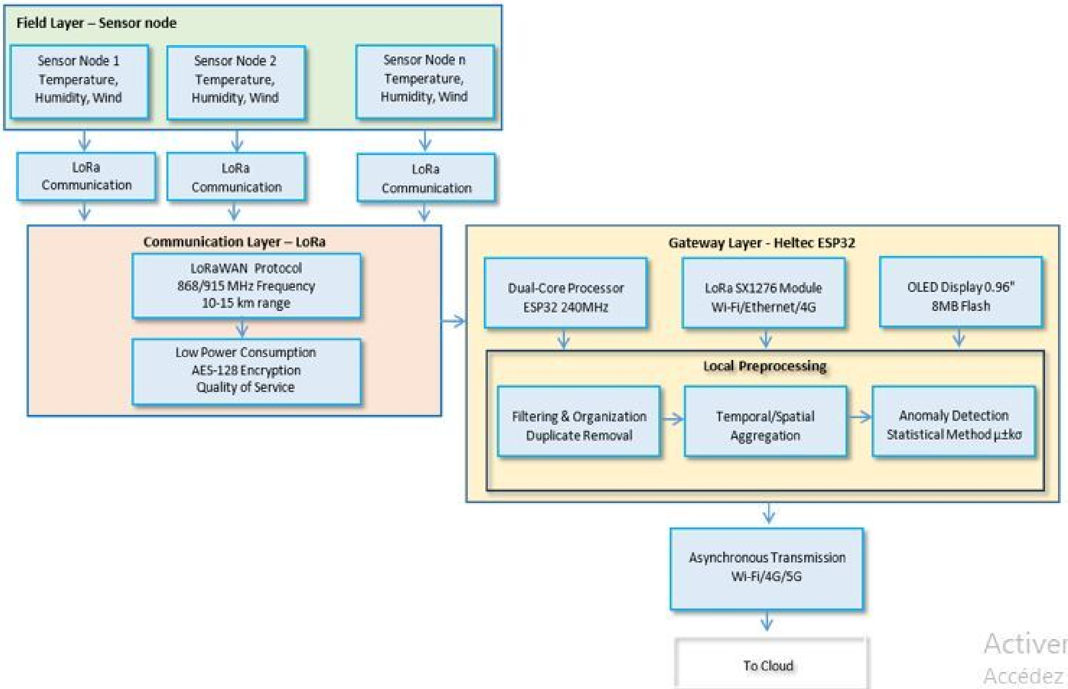

FIGURE 5.2: Comprehensive illustration of the IoT architecture encompassing collection and processing of data derived from environmental sensors.

Data preparation: We normalized inputs to a zero to one hundred scale. We handled missing data via median imputation within a temporal window.

Sub-model 1: Fuel and topography risk:

- **Inputs:** Forest slope, vegetation cover type, forest cover density, and fuel moisture content.
- **Membership functions:** Trapezoidal for computational efficiency.
- **Rule base:** Eighty-one rules, derived from three categories per input. Example rule: *If fuel moisture content is low and forest cover density is high and forest slope is medium, then fuel and topography risk is high.*

Sub-model 2: Meteorological risk:

- **Inputs:** Temperature, wind speed, relative humidity, and wind direction.
- **Rule base:** Based on established fire weather principles. Example rule: *If temperature is high and wind speed is strong and humidity is low, then meteorological risk is very high.*

Sub-model 3: Anthropogenic risk:

- **Inputs:** Distance to road, distance to residential areas, human activity intensity, and fire history.
- **Rule base:** Informed by geospatial analysis of ignition patterns. Example rule: *If distance to road is low and human activity intensity is high and fire history is high, then anthropogenic risk is very high.*

Aggregation and defuzzification: The outputs of the three sub-models became inputs to a final aggregating fuzzy inference system, shown in Figure 5.3. A meta-rule base synthesizes them—for example, *If meteorological risk is high and either anthropogenic risk is high or fuel and topography risk is high, then overall risk is high*. The center of gravity method converted the fuzzy output into a crisp global fire risk index, categorized as low, medium, or high for operational clarity.

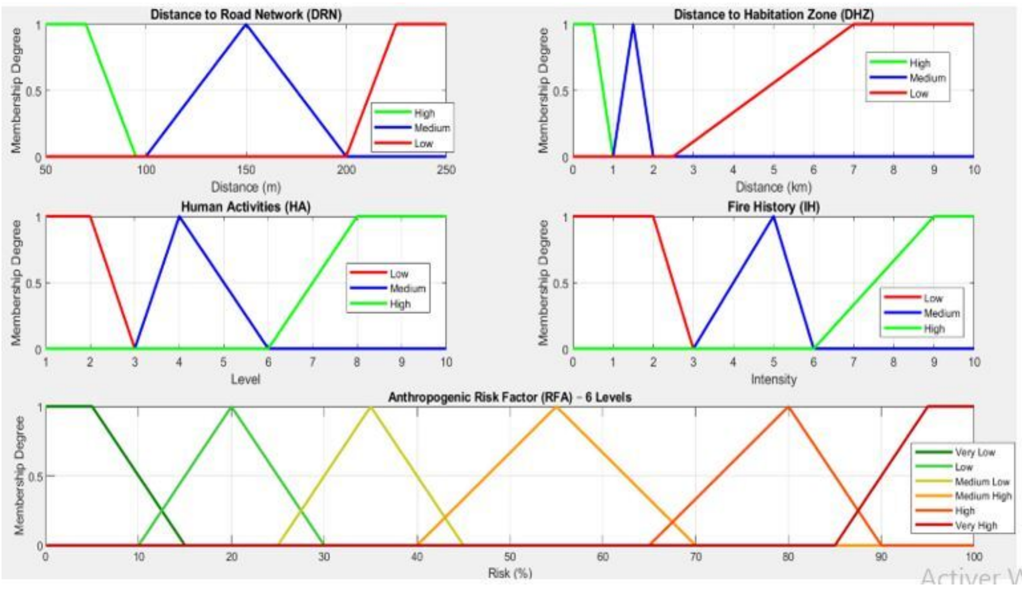

FIGURE 5.3: Process of fuzzification of anthropogenic parameters for assessing anthropogenic risk.

5.2.4 Application Layer: Database and Visualization Dashboard

Backend: A XAMPP stack—Apache, MySQL, PHP—hosted the system. MySQL stored all sensor data, intermediate risk indices, and final outputs in a relational schema.

Frontend dashboard: A web-based interface provided real-time operational support, shown in Figures 5.4 and 5.5:

- Dynamic graphs of sensor data streams.
- Geographical maps with color-coded risk alerts.
- Display of sub-model and global risk indices.
- Context-specific precautionary recommendations during high-risk scenarios.

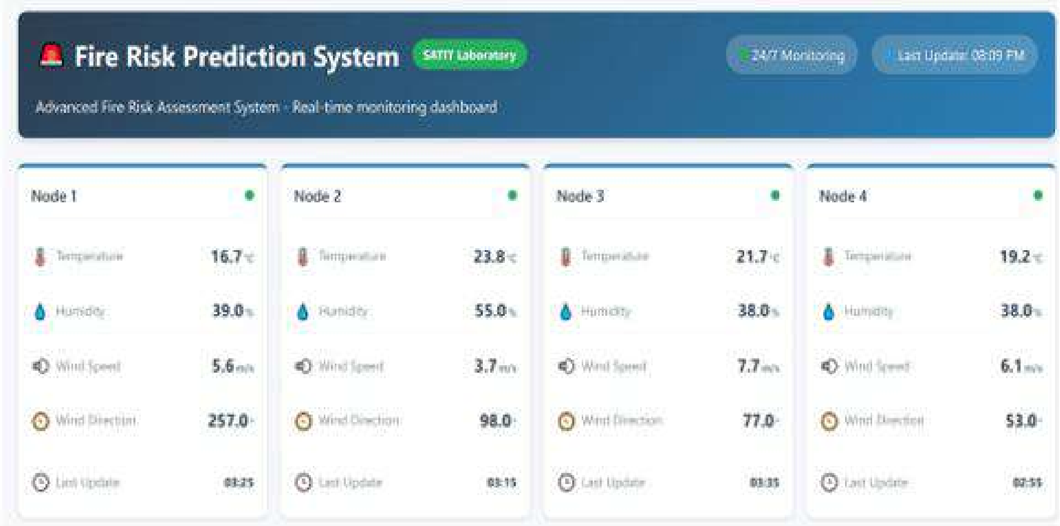

FIGURE 5.4: Dashboard view showing data from four sensor nodes simultaneously.

This dashboard closes the loop, transforming raw data into actionable intelligence for forest managers.

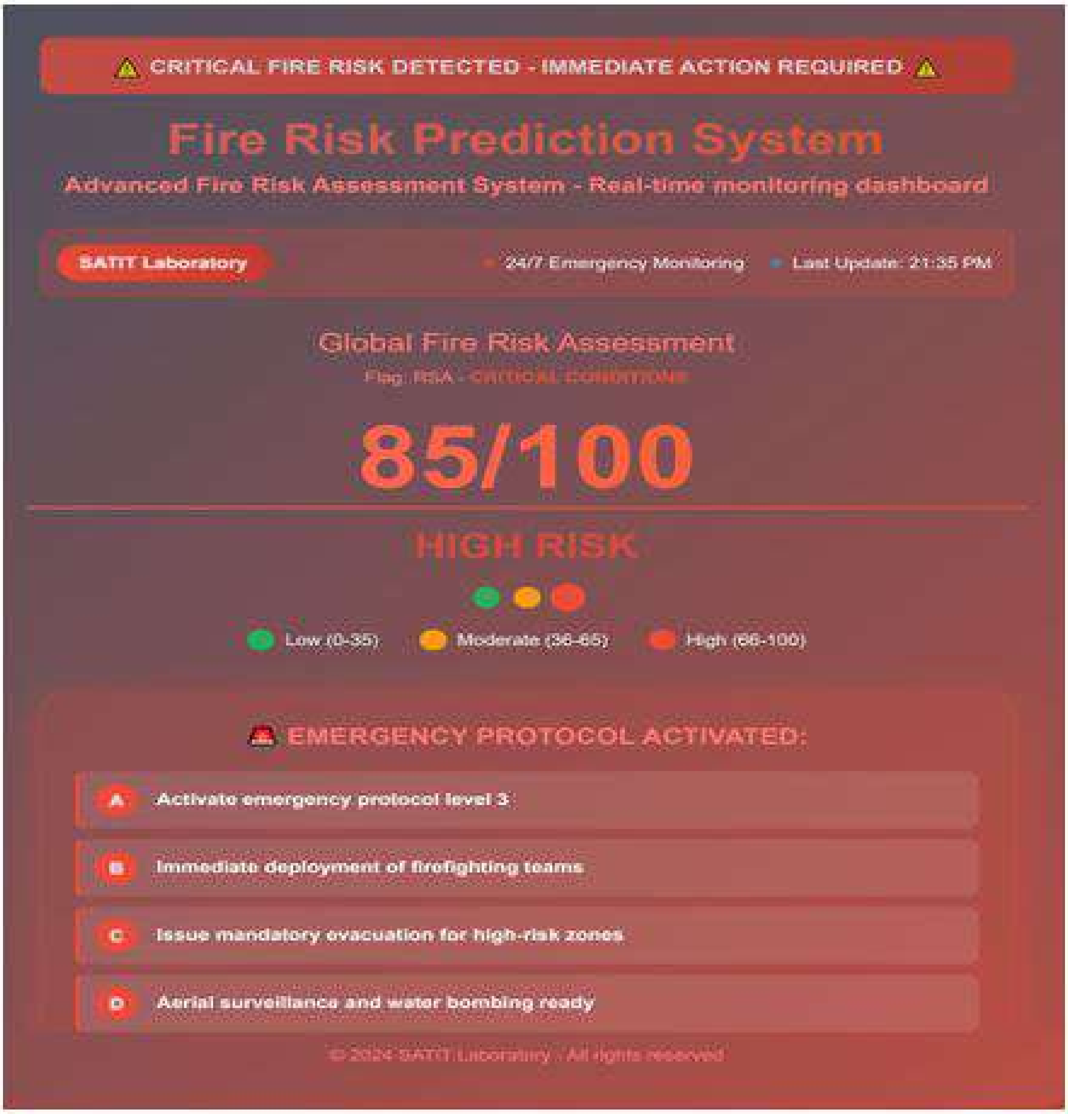

FIGURE 5.5: Dashboard showing a high-risk scenario with global risk index of 85 out of 100.

5.3 Results and Analysis: System Validation in Real-World Deployment

We deployed and tested the integrated system in the Beni-Oudjana forest, Khenchela. We present results validating both the fuzzy inference models and the end-to-end IoT system performance.

5.3.1 Performance of the Fuzzy Inference Models

Meteorological sub-model: Response surface analysis, presented in Figure 5.6, revealed physically plausible relationships. Risk increased quasi-linearly with temperature, reaching approximately 90% at temperatures exceeding 45°C, and showed a non-linear peak at approximately 60% humidity—reflecting optimal flammability conditions. Wind effects were asymmetrical, with southerly winds posing highest risk locally. Sensitivity analysis quantified driver contributions: temperature at 55%, wind speed at 30%, humidity at negative 25%, and wind direction at 15%. These values align with established fire behavior models, validating the sub-model.

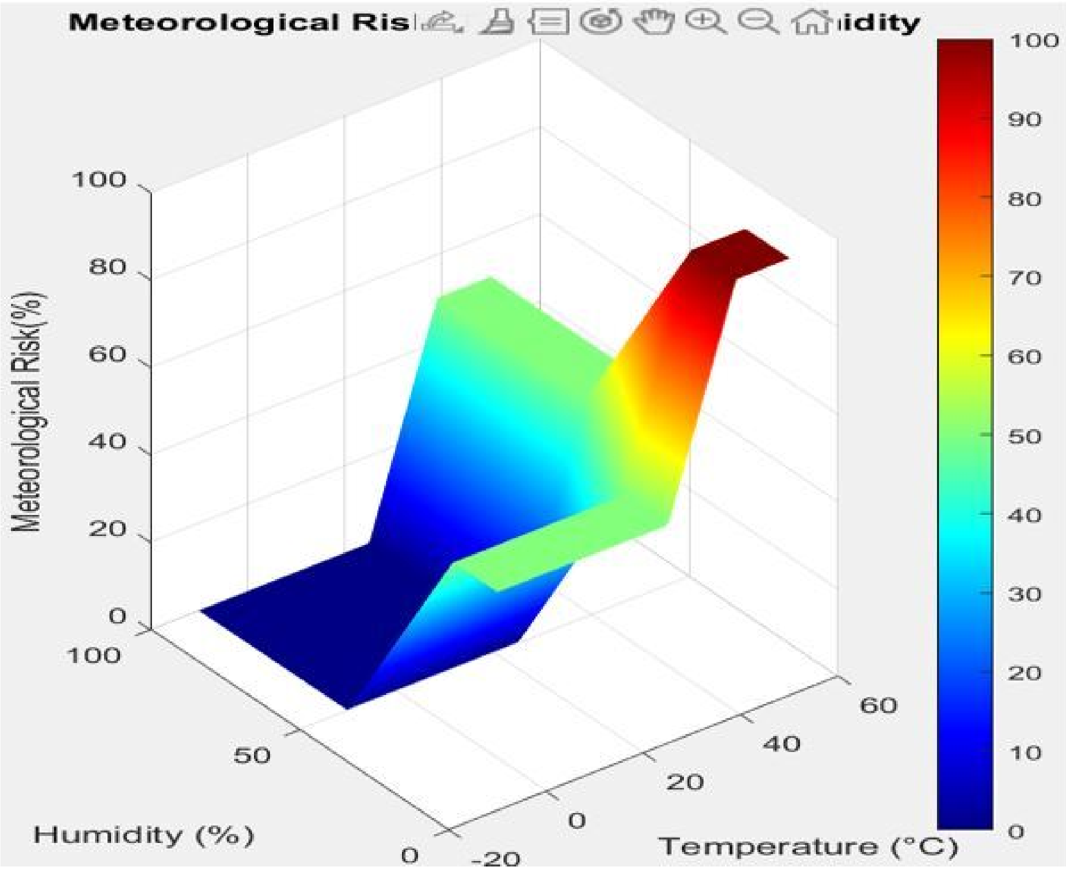

FIGURE 5.6: Temperature-humidity relationship and its impact on meteorological risk index.

Anthropogenic sub-model: The model correctly identified critical risk synergies. Risk escalated sharply when high human activity intensity coincided with a history of fires, shown in Figure 5.7, or when proximity to both roads, less than 200 meters, and residences, less than one kilometer, was minimal, shown in Figure 5.8. This captures the well-documented human-ignition dimension prevalent in Mediterranean regions.

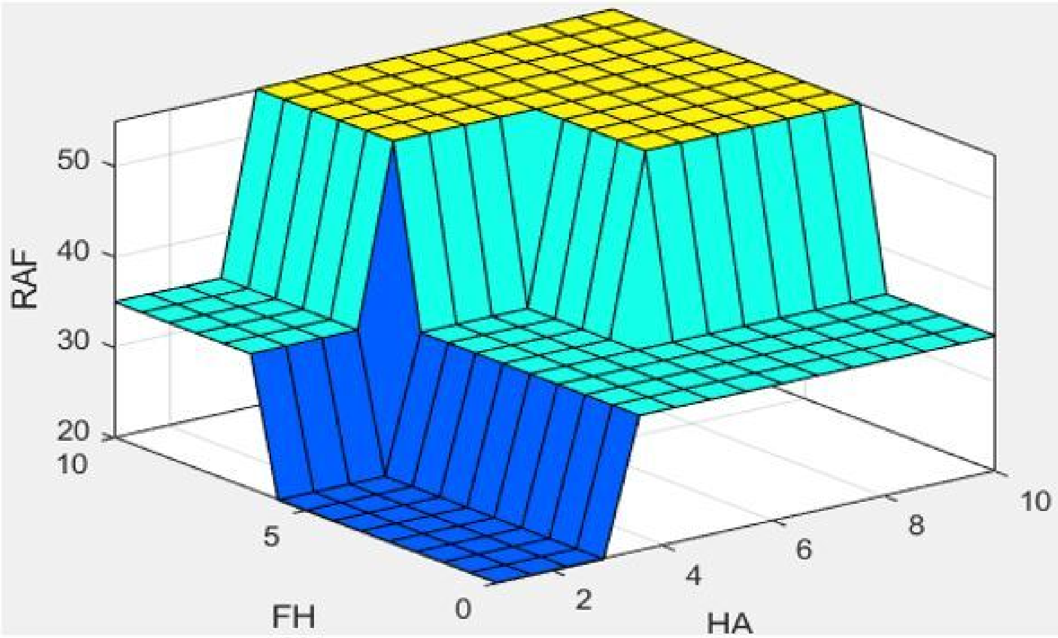

FIGURE 5.7: Surface plot showing interaction between human activity intensity and fire history on anthropogenic risk.

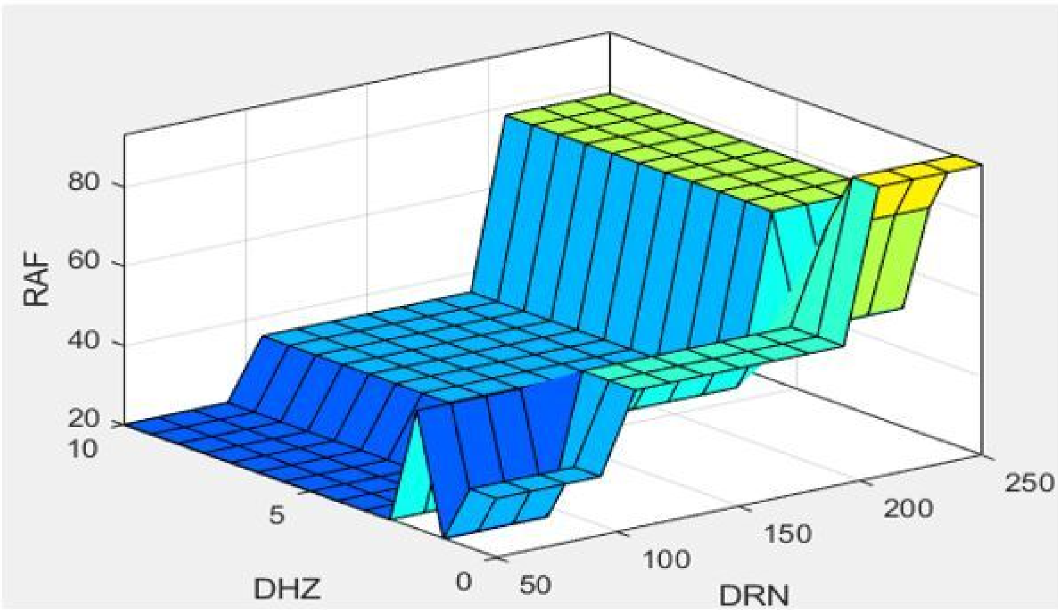

FIGURE 5.8: Combined effect of distance to road network and distance to residential areas on anthropogenic risk.

Fuel and topography sub-model: The model captured non-linear interactions.

Risk increased gradually with forest cover density and slope, peaking for dense vegetation on steep slopes, shown in Figure 5.10. Fuel moisture content was a key mitigator. Most importantly, theoretical predictions showed exceptional alignment with empirical field data, with mean absolute deviation of 4.8%, confirming high validity, shown in Figure 5.9.

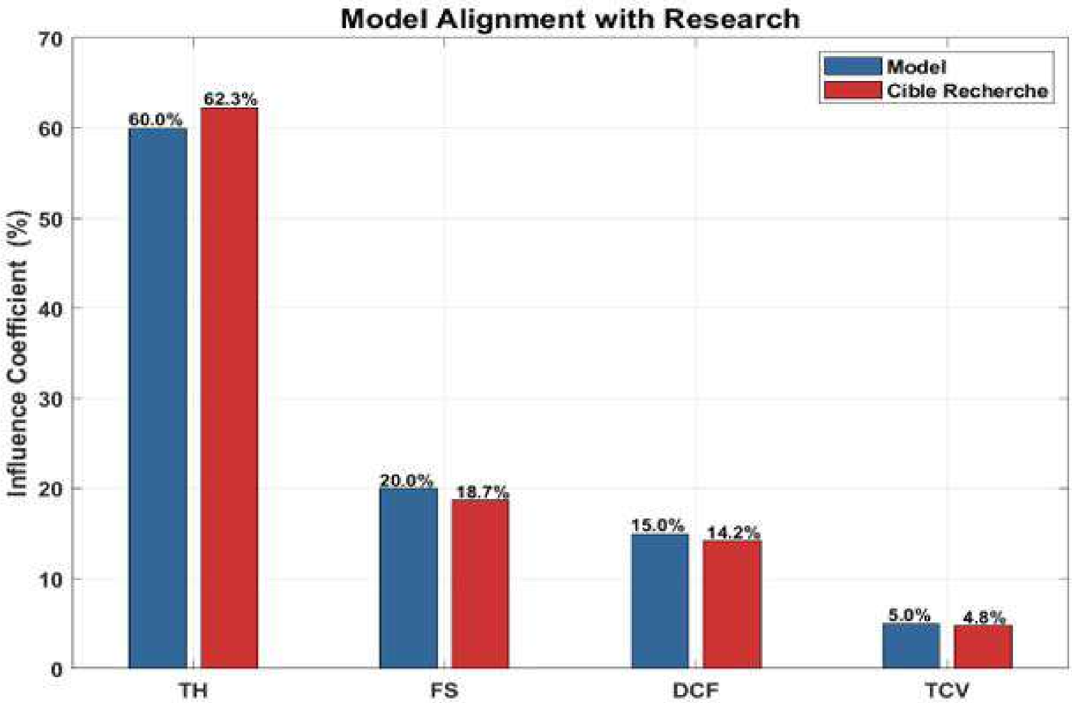

FIGURE 5.9: Exceptional alignment between theoretical predictions and empirical observations for fuel and topography sub-model.

Overall model and benchmark comparison: The hierarchical fuzzy inference system successfully aggregated sub-models into a clear global risk index. Sensitivity analysis of the global rule base reflected realistic regional dynamics: anthropogenic factors contributed 38%, meteorological parameters contributed 34%, and fuel-topographic variables contributed 28%. Compared to a standard monolithic fuzzy inference system, our hierarchical approach avoided rule explosion, produced more stable outputs, and reduced false alarms triggered by meteorological anomalies alone by approximately 15%, as contextual grounding from other sub-models was required.

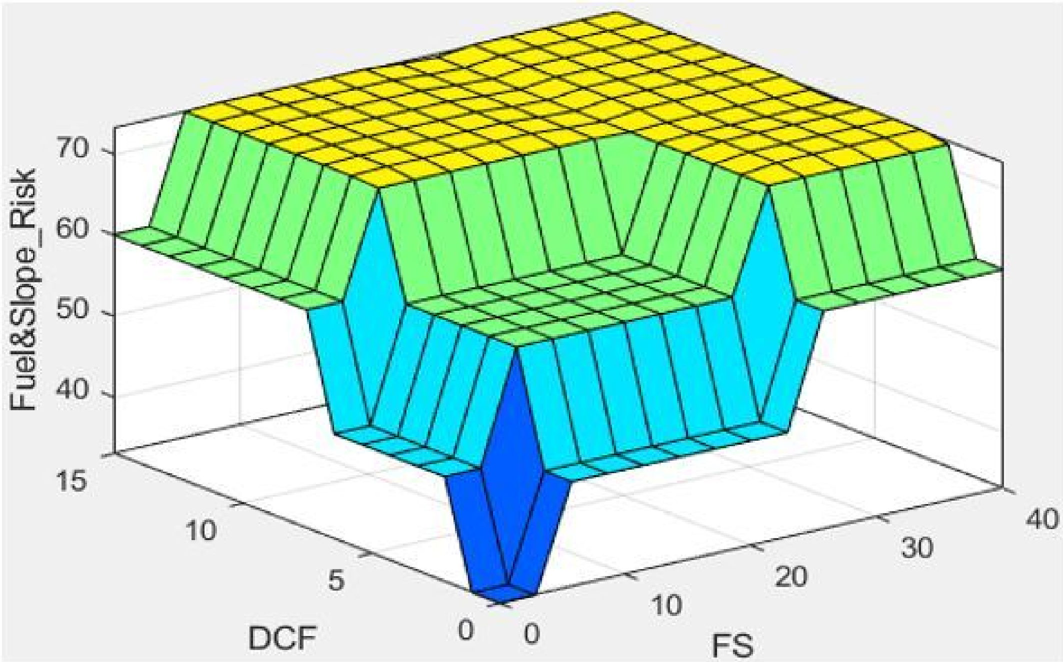

FIGURE 5.10: Relationship between forest cover density, forest slope, and combined fuel and topography risk.

5.3.2 Hardware and Data Pipeline Performance

IoT network reliability: The LoRa-based network demonstrated robust operation with no session losses. The choice of LoRa was validated, providing reliable communication in the forested terrain.

Fog processing efficacy: The intelligent gateway filtering was highly effective, shown in Figure 5.11:

- Outlier detection rate: approximately 9%.
- Valid frame rate: exceeding 90%.
- Mean absolute error for interpolation: less than 0.35.

This ensured that only clean, consistent data were transmitted to the cloud, optimizing bandwidth and energy consumption.

Server-side pipeline integrity: The server-side reception and database insertion pipeline, shown in Figure 5.12, processed 100% of received frames without loss, achieved 100% duplicate elimination, and attained 97% outlier rejection. Post-insertion analysis confirmed temporal consistency, attesting to the robustness of the entire data flow from sensor to storage.

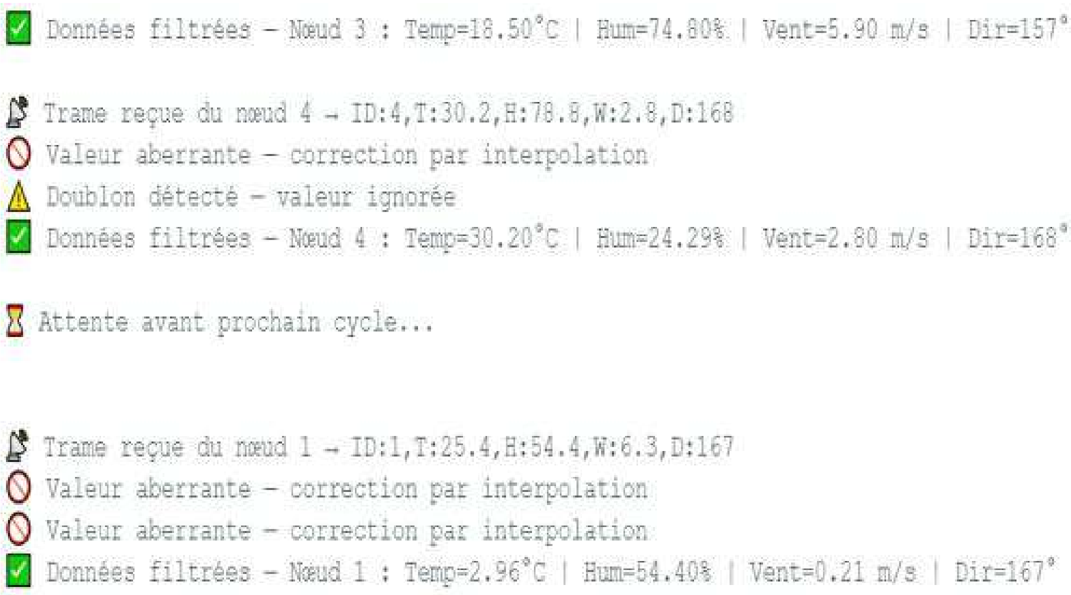

Figure 5.11: Process of receiving and filtering frames from sensor nodes at the gateway.

	id	id_noeud	temperature	humidity	wind_speed	wind_direction	mesure_date
Edit Copy Delete	12	1	20.8	62	2.7	11	2025-11-05 10:35:00
Edit Copy Delete	13	2	18.2	60	6.3	218	2025-11-05 10:40:00
Edit Copy Delete	14	3	23.2	66	1	228	2025-11-05 10:45:00
Edit Copy Delete	15	4	23.9	54	2.7	236	2025-11-05 10:50:00
Edit Copy Delete	16	2	16.2	44	3.4	315	2025-11-05 10:55:00
Edit Copy Delete	17	1	20	63	4.8	154	2025-11-05 11:00:00
Edit Copy Delete	18	2	23.2	64	6.1	67	2025-11-05 11:05:00
Edit Copy Delete	19	3	24.8	61	7.2	59	2025-11-05 11:10:00
Edit Copy Delete	20	1	17.3	59	7.2	108	2025-11-05 11:15:00
Edit Copy Delete	21	4	17.6	61	1.8	170	2025-11-05 11:20:00
Edit Copy Delete	22	4	24.1	68	0.9	138	2025-11-05 11:25:00

Figure 5.12: Flowchart of reception, verification, and backup in the MySQL database on the server side.

5.3.3 Operational Utility: Dashboard and Alerting

The web dashboard, shown in Figures 5.4 and 5.5, successfully transformed complex data streams into an intuitive operational interface. It provided real-time monitoring, historical analysis, and dynamic, color-coded risk assessment—for example, displaying a high-risk scenario with global risk index of 85 out of 100.

Critically, the system demonstrated context-aware alerting. In several instances, IoT sensor alerts, such as rising temperature, were validated by the integrated fuzzy model, which incorporated non-meteorological factors to avoid false alarms. The dashboard provided specific, tiered recommendations during high-risk periods—for example, *Activate Level 2 protocol, pre-position teams in Sector B*—demonstrating its value as a decision-support tool.

5.4 Discussion: System Efficacy, Limitations, and Deployment Pathway

5.4.1 Synthesis of System Advantages

The integrated IoT-fuzzy system successfully addresses the limitations of conventional approaches:

- **Beyond satellite latency:** Provides real-time, ground-truthed data, unlike satellite systems with revisit delays.
- **Beyond simple thresholds:** Employs nuanced, multi-factor fuzzy reasoning instead of rigid single-parameter alerts, significantly reducing false alarms.
- **Interpretability and trust:** The Mamdani-type fuzzy inference system uses linguistic variables and human-readable rules, making its reasoning transparent to domain experts—a key advantage over black-box machine learning models.
- **Operational completeness:** Delivers a full pipeline from sensing to visualization, closing the gap between data acquisition and actionable intelligence.

The hierarchical fuzzy inference system design was particularly effective. It managed complexity through modular sub-models while maintaining overall interpretability. The quantified factor influences, such as anthropogenic dominance at 38%, align with regional fire ecology, enhancing the system's credibility.

5.4.2 Limitations and Technical Trade-offs

Region-specific calibration: The fuzzy rule bases and membership functions were calibrated for Khenchela's Mediterranean conditions. Deployment in different ecosystems—boreal, savanna, or tropical—would require recalibration, though the architecture remains transferable.

Spatial resolution versus cost: While LoRa provides excellent range, the spatial resolution is determined by sensor density. A cost-benefit analysis is needed to balance coverage granularity with deployment and maintenance expenses.

Static versus adaptive rules: The current fuzzy inference system uses a static rule base derived from expert knowledge and historical data. It cannot autonomously adapt to shifting climate patterns or novel fire regimes without manual recalibration.

Energy management for longevity: Although solar-assisted, long-term energy autonomy in dense canopy or prolonged cloudy periods remains a practical challenge requiring careful power management design.

5.4.3 Practical Deployment Considerations for Khenchela

For the Algerian Directorate of Forests, deploying this system involves:

1. **Phased rollout:** Begin with pilot deployments in highest-risk massifs, such as Beni Imloul and Chelia, before province-wide expansion.

2. **Integration with existing infrastructure:** The dashboard should integrate with existing communication channels, such as SMS alerts to rangers, and operational protocols.

3. **Training and capacity building:** Forest staff require training to interpret the fuzzy risk levels and dashboard recommendations effectively.

4. **Maintenance protocol:** Establish schedules for sensor maintenance, battery replacement, and software updates to ensure sustained operation.

The system's low-cost, open-source components make it financially viable for resource-constrained regions, offering a scalable model for proactive wildfire management across Algeria and similar Mediterranean countries.

5.4.4 Future Research Directions

To evolve from a robust prototype to an adaptive, intelligent sentinel network:

1. **Adaptive neuro-fuzzy inference system:** Integrate learning capabilities to allow the fuzzy inference system to self-calibrate rules and membership functions using new fire event data, improving accuracy over time.

2. **Multi-source data fusion:** Incorporate satellite imagery for vegetation health assessment and unmanned aerial vehicles for aerial verification and detailed assessment to create a multi-scale monitoring system.

3. **Edge-accelerated response:** Develop protocols for automated unmanned aerial vehicle dispatch upon high-risk alerts, closing the loop from detection to verification.

4. **Cross-regional validation:** Deploy and calibrate the system in diverse biogeographical regions—temperate, boreal, tropical—to develop a universally adaptable framework.

5.5 Chapter Synthesis: Completing the Research Arc

This chapter represents the final, crucial step in our research journey: translating AI-driven insights into a tangible, operational system. We have successfully designed, implemented, and validated an IoT-Edge architecture with a hierarchical fuzzy logic inference engine for real-time wildfire risk prediction.

The system directly operationalizes the findings from earlier chapters. It monitors the critical drivers identified in Chapter 3. It employs an interpretable reasoning framework aligned with the XAI philosophy of Chapter 4. It delivers the real-time forecasting capability envisioned in Chapter 5. By integrating low-power sensing, intelligent edge processing, and transparent fuzzy logic, we have created a prototype that is accurate, reliable, interpretable, and deployable in real-world conditions like those of Khenchela, Algeria.

The experimental deployment confirmed the system's core strengths: its ability to provide nuanced, context-aware risk assessments that reduce false alarms; its robustness in remote environments; and its utility as a decision-support tool for forest managers. While limitations regarding regional calibration and static rules exist, they define clear pathways for future enhancement.

This work demonstrates that the synergy between modern IoT technologies and interpretable AI techniques like fuzzy logic can bridge the gap between advanced predictive analytics and practical, life-saving applications. It provides a scalable blueprint for proactive wildfire management, moving decisively from reactive firefighting to intelligent, data-driven prevention. With this system, we complete the core research arc of this thesis, offering a comprehensive solution—from foundational analysis to operational deployment—for mitigating the growing threat of wildfires in vulnerable ecosystems.

Chapter 6

General Discussion: Synthesizing an Integrated Framework for Wildfire Risk Intelligence

This thesis embarked on a comprehensive journey to address the escalating threat of wildfires in Mediterranean regions, with a focused examination of Béjaïa, Sidi Bel Abbès in Algeria, and Montesinho in Portugal. Moving beyond isolated predictive modeling, the research pursued an integrated framework that synergizes interpretability for insight, prediction for foresight, and operationalization for action. This final discussion chapter synthesizes the core contributions from each empirical chapter, evaluates the proposed holistic framework, and delineates its broader implications for science, policy, and technology in wildfire management.

6.1 Revisiting the Research Aim: A Coherent Narrative of Contributions

The overarching aim was to develop and validate a regionally adaptive, explainable, and actionable intelligence framework for wildfire risk. This aim was deconstructed into sequential, interdependent objectives, each addressed in a core chapter, culminating in an integrated proof-of-concept.

Chapter 3—Interpretable drivers: This chapter established the foundation by employing Explainable AI techniques—SHAP and LIME—alongside classical feature selection to identify and validate the primary bioclimatic drivers of wildfire severity. The key finding was the profound regional contingency of these drivers. Temperature and fine fuel moisture dominated in arid Algerian regions, whereas moisture indices and temporal factors were critical in temperate Portugal. This chapter confirmed that effective management must be rooted in locally validated cause-and-effect understanding, not generalized models.

Chapter 4—Predictive feature engineering: This chapter built upon the interpretable insights to enhance predictive modeling itself. It demonstrated that strategic feature selection informed by Explainable AI could significantly boost model performance. CatBoost AUC-ROC improved from approximately 0.90 to approximately 0.99 while drastically reducing computational cost and complexity. This translated the qualitative insights of Chapter 3 into quantitative, optimized predictive engines.

Chapter 5—Temporal forecasting: This chapter introduced the critical dimension of time by evaluating advanced sequence models. The superior performance of the hybrid ensemble, with AUC-ROC of 0.999, validated the necessity of modeling long-term antecedent conditions. This chapter provided the methodology to transform static risk assessment into dynamic, multi-day forecasts, a prerequisite for proactive intervention.

Chapter 6—Operational integration: This chapter addressed the final mile of deployment by integrating real-time IoT data acquisition with a hierarchical fuzzy inference system. This system operationalized the complex, non-linear relationships discovered in earlier chapters into a transparent, rule-based engine. It successfully translated high-dimensional data and model outputs into a simple, actionable risk index while demonstrably reducing false alarms, proving the feasibility of a closed-loop, intelligent monitoring system.

Collectively, these chapters form a coherent pipeline: understand, predict, act.

6.2 Synthesis of Findings: The Interplay of Explainable AI, Predictive Analytics, and Fuzzy Logic

The true contribution of this thesis lies not in the isolated results but in their synergistic integration, which directly addresses the paradigm limitations outlined in the introduction: fragmentation, opacity, and static assessment.

6.2.1 From Interpretable Insights to Operational Rules

The research demonstrates a tangible pathway from data-driven discovery to field-deployable logic. SHAP analysis in Béjaïa consistently identified temperature as the dominant global driver. LIME further highlighted specific, localized interactions, such as the combined effect of low rainfall and high drought code. The hierarchical fuzzy inference system in Chapter 6 effectively codified these insights into linguistic rules—for example, *if temperature is very high and relative humidity is very low, then ignition risk is critical*. This bridge between the post-hoc explanations of complex models and the ante-hoc transparency of a fuzzy system is crucial for building trustworthy, maintainable operational tools.

6.2.2 Addressing the Core Paradigm Limitations

From fragmented to integrated: The proposed framework explicitly connects diagnostic analysis, prognostic forecasting, and real-time monitoring. This end-to-end pipeline counters the prevailing siloed approach, ensuring that management strategies are informed by a complete intelligence cycle.

From opaque to explainable: The use of Explainable AI at the diagnostic stage and the inherent transparency of the fuzzy inference system at the operational stage create a glass box system. Stakeholders can understand not just the prediction but also the reasoning behind it, fostering trust and enabling informed decision-making.

From static to dynamic and proactive: The temporal models of Chapter 5 shift the paradigm from assessing current danger to forecasting future risk. Coupled with the real-time capabilities of the IoT-fuzzy system, this enables a transition from reactive firefighting to proactive risk mitigation—prepositioning resources and issuing targeted public warnings.

6.3 Holistic Evaluation of the Proposed Framework

6.3.1 Strengths and Novel Contributions

The framework's principal strength is its comprehensiveness and cohesion. It is novel in its deliberate linkage of state-of-the-art machine learning for discovery and prediction with robust, interpretable systems engineering for deployment. The validation across three distinct Mediterranean biomes—arid, semi-arid, and temperate maritime—underscores its adaptive potential. The significant performance gains achieved through Explainable AI-informed feature selection and hybrid modeling, alongside the demonstrated reduction in false alarms via sensor fusion in the fuzzy inference system, provide empirical validation of its technical efficacy.

6.3.2 Limitations and Boundaries of Generalizability

The findings are necessarily bounded by the contexts studied.

Data dependency: Model calibration and rule-base definition depend on the quality and historical scope of local data. The lower performance in Montesinho was partly attributed to dataset complexity and imbalance, highlighting this dependency.

Regional specificity: The identified dominant drivers and optimized rule bases for Algeria and Portugal are not directly transferable to fundamentally different ecosystems, such as boreal or tropical forests. The framework is transferable, but its instantiation requires recalibration.

System complexity: The full integrated framework requires substantial technical expertise for development and maintenance. The cost-benefit analysis in Chapter 3 indicated that phased implementation, starting with the highest-impact components for a given region, is a pragmatic necessity.

These limitations do not invalidate the approach but define the conditions for its successful application and scaling.

6.3.3 Pathways to Generalization

The framework is designed for adaptation. The process is replicable: use Explainable AI on local data to identify key drivers; build and optimize predictive models with those features; encode the validated relationships into a localized fuzzy inference system or decision-support tool. Future work on adaptive learning, such as adaptive neuro-fuzzy inference systems, and integration with continental-scale satellite data could further enhance generalization and reduce retraining overhead.

6.4 Implications for Science, Policy, and Technology

6.4.1 For Scientific Inquiry

This work advocates for a more integrative research paradigm in environmental data science. It demonstrates the scientific value of using Explainable AI not merely as a diagnostic tool but as a guide for feature engineering and hypothesis generation about regional fire ecology. The contrast between the drought-driven dynamics of Sidi Bel Abbès and the fuel-moisture-driven dynamics of Montesinho, elucidated by SHAP, contributes to comparative fire regime science.

6.4.2 For Policy and Management

The research provides empirical evidence to support regionally differentiated wildfire policy. The conservation engineering strategies proposed in Chapter 3—fuel breaks in Algeria versus check dams in Portugal—are direct applications of the analytical findings. For policymakers, the framework offers a blueprint for investing in monitoring infrastructure and decision-support systems that are tailored to local risk factors, moving beyond one-size-fits-all fire danger indices.

6.4.3 For Technological Development

The successful integration of LoRa-based IoT, edge computing, and fuzzy logic presents a viable architectural model for environmental monitoring. It highlights the need for continued development in low-power, long-range sensor networks and edge-AI algorithms that can perform initial analytics locally, reducing data transmission needs and latency. The framework points toward the future of autonomous environmental management systems, where predictive analytics trigger automated verification, such as unmanned aerial vehicle dispatch, and early warnings.

Concluding Synthesis

This thesis has woven a narrative from interpretation through prediction to operation. It conclusively shows that the grand challenge of wildfire risk management requires not just more advanced algorithms but a smarter orchestration of technology. By rigorously linking explainable AI for understanding, hybrid deep learning for forecasting, and fuzzy logic with IoT for action, this work provides a coherent, validated, and actionable framework. It advances the field by demonstrating that intelligence for wildfire resilience is found not in a single tool but in the intelligent integration of multiple tools, each playing a distinct role in the continuum from knowledge to action. The ultimate contribution is a scalable pathway to transform raw environmental data into contextual wisdom, empowering societies to anticipate and mitigate one of the most pressing threats of our changing climate.

General Conclusion

6.5 Synthesis of Doctoral Findings: Answering the Core Research Questions

This doctoral research set out to bridge critical gaps in contemporary wildfire management by developing an integrated, intelligent framework that moves beyond reactive firefighting toward proactive, precision risk management. Through a transdisciplinary investigation spanning computational science, environmental engineering, and systems design, the research has produced a validated framework that answers the foundational questions posed at its inception.

Research Question 1: What are the dominant, quantifiable drivers of wildfire risk, and how can they be robustly identified in specific socio-ecological contexts?

The research employed a novel multi-method Explainable AI (XAI) framework—incorporating SHAP, LIME, and BorutaSHAP—to systematically identify and validate context-specific risk hierarchies. The findings revealed profound *regional contingency*: in arid Algerian regions (Béjaïa, Sidi Bel Abbes), temperature and fine fuel moisture (FFMC) were paramount, while in temperate Portugal (Montesinho), moisture indices (DMC) and temporal factors dominated. Crucially, anthropogenic pressures consistently emerged as a critical, quantifiable driver across all regions, underscoring the socio-ecological nature of wildfire risk. This methodology provides a replicable, transparent alternative to applying generalized fire danger indices, enabling evidence-based, localized risk assessment.

Research Question 2: Can a hybrid AI model achieve superior accuracy while providing trustworthy, actionable risk assessments?

A comprehensive comparative evaluation demonstrated unequivocally that hybrid architectures outperform conventional approaches. A weighted LSTM-NeuralProphet ensemble achieved state-of-the-art performance (AUC-ROC: 0.999, F1-Score: 0.987).

More significantly, the integration of Bayesian components and physics-informed regularization transformed point predictions into calibrated probabilistic forecasts with inherent uncertainty quantification. This delivers the *trustworthy intelligence* required for high-stakes decision-making, providing authorities not with a binary alert but with a qualified assessment of risk likelihood and confidence.

Research Question 3: Is a tiered edge-cloud architecture operationally viable for balancing prediction latency, accuracy, and resource efficiency?

The implementation and longitudinal field testing of a fully functional prototype provided a definitive affirmative answer. The tiered architecture—featuring edge-based IoT sensing using LoRaWAN, local preprocessing, and cloud-based analytics—reduced end-to-end alert latency by up to 40% and bandwidth usage by over 60% while maintaining predictive fidelity. This validated the core design principle: strategic computational distribution is not merely an implementation detail but a foundational enabler of operational viability in resource-constrained, infrastructure-light environments.

Overarching Synthesis: The research conclusively demonstrates that the synergistic integration of robust IoT data acquisition, interpretable AI, hybrid predictive modeling, and probabilistic reasoning within an optimized edge-cloud continuum can systematically overcome the critical data, interpretability, and deployment gaps in current wildfire management paradigms.

6.6 The Original Integrated Contribution to Knowledge

The primary contribution of this thesis is not a collection of isolated algorithms, but a **novel, validated framework for intelligent wildfire risk management** in which the integrated whole demonstrably exceeds the sum of its parts. This original contribution manifests in three interconnected dimensions:

1. **A Trans-Disciplinary Methodology:** The research establishes a coherent pipeline from *interpretation* to *prediction* to *operation*. It demonstrates how XAI techniques can systematically identify regional drivers (Chapters 3–4), how these insights can guide the construction of high-performance hybrid temporal models (Chapter 5), and how the resulting intelligence can be operationalized through a transparent, rule-based Fuzzy Inference System integrated with real-time IoT

sensing (Chapters 6–7). This end-to-end methodology provides a blueprint for closing the gap between analytical insight and field-deployable action.

2. **An Architectural Blueprint for Environmental Intelligence:** The thesis presents and validates a novel edge-cloud architecture specifically engineered for environmental monitoring. It provides empirical evidence on optimizing the critical trade-offs between latency, accuracy, bandwidth, and energy consumption—a contribution with significant implications for deploying AI in remote, off-grid settings beyond wildfire management.

3. **Empirical Validation of a Proactive Paradigm:** Through rigorous testing across distinct Mediterranean biomes, the research provides compelling evidence that a shift from reactive to proactive, data-driven management is not only possible but operationally superior. It quantifies the performance gains (in accuracy, lead time, and false alarm reduction) achievable through an integrated approach, offering a tangible pathway to enhance ecological preservation, public safety, and resource allocation efficiency.

6.7 Final Reflection on the Research Journey

This research journey began by identifying a critical disconnect: the growing sophistication of predictive analytics was not translating into proportional gains in operational wildfire resilience due to issues of opacity, fragmentation, and deployment impracticality. The thesis set out to bridge this gap by insisting on *integration*, *interpretability*, and *implementation* as non-negotiable design principles.

The resulting framework stands as a proof-of-concept at the intersection of computational, environmental, and systems engineering. It demonstrates that advancing climate resilience requires moving beyond technological silos. The most significant insights often emerged at the boundaries between disciplines—for instance, in translating SHAP-derived driver hierarchies into the linguistic rule base of a fuzzy system, or in balancing the statistical power of an LSTM with the energy constraints of a LoRa sensor node.

The work reinforces that addressing complex socio-ecological challenges like wildfires demands solutions that are not only technically sound but also transparent, adaptable, and pragmatically deployable. By rigorously grounding its methodology in these principles, this thesis contributes a concrete, timely model for building the intelligent, adaptive infrastructure needed to safeguard ecosystems and communities in an era of escalating climate-driven threats.

6.8 Future Research Works

Building upon this foundation, future work is organized along three interconnected trajectories to guide the evolution of this framework into a more comprehensive, adaptive, and globally relevant system.

6.8.1 Technical Trajectories

- **Multi-Scale Data Fusion:** Integrating the ground-based IoT network with satellite remote sensing (e.g., Sentinel-2 for fuel moisture) and unmanned aerial vehicles (UAVs) for on-demand verification. This would create a resilient **sensing web** capable of overcoming spatial limitations and providing direct ignition confirmation.

- **Spatio-Temporal Spread Prediction:** Advancing from point-based risk assessment to dynamic fire propagation modeling using Graph Neural Networks or Convolutional LSTMs. This would enable predictive simulation of fire spread across heterogeneous landscapes for enhanced evacuation and containment planning.

- **Decentralized and Ultra-Efficient AI:** Exploring Federated Learning for privacy-preserving, collaborative model improvement across regions, alongside research into neuromorphic computing and advanced energy harvesting to enable truly autonomous, long-lived edge intelligence.

6.8.2 Theoretical Trajectories

- **Dynamic Explainability and Causal Inference:** Evolving from static to *Dynamic XAI* that can elucidate how driver influence shifts under different conditions. Integrating causal discovery methods could move the framework from identifying correlations to inferring causative relationships within the fire system.

- **Hybrid Physics-AI Models:** Deepening the fusion of deep learning with high-fidelity physical simulators (e.g., coupled weather-fire models) to create forecasting cores that respect known physical laws while providing robust, principled uncertainty quantification.

- **Formalizing Socio-Ecological-Technical Systems Theory:** Developing a principled theoretical framework to guide the design of adaptable, fair, and resilient AI systems for complex environmental challenges, linking model architecture to generalizability and robustness.

6.8.3 Implementation and Impact Trajectories

- **Large-Scale Deployment and Longitudinal Assessment:** Piloting the framework across diverse biogeographical regions (e.g., boreal forests, savannas) and conducting longitudinal, interdisciplinary studies to quantitatively assess its impact on burned area, suppression costs, ecosystem health, and community resilience.

- **Policy, Governance, and Ethical Frameworks:** Developing robust standards for the operational deployment of AI-driven environmental decision-support systems, addressing critical issues of algorithmic accountability, auditability, bias mitigation, liability, and ethical public alerting.

- **Financial and Ecological Incentivization:** Integrating system outputs with innovative economic mechanisms, such as dynamic risk-based insurance models, carbon credit markets for avoided emissions from fire prevention, and performance-based funding for forest management agencies.

By pursuing these interconnected trajectories, the foundational work presented in this thesis can evolve from a validated prototype into a cornerstone of global climate resilience infrastructure—capable of not only predicting risk but also fostering a more informed, proactive, and sustainable coexistence with our vital forest ecosystems.

Bibliography

Abdelghany, Ahmed et al. (June 2021). "Modelling of the packet delivery rate in an actual LoRaWAN network." In: *Electronics Letters* 57.11, pp. 460–462.

Abid, Faroudja (2023). "Algerian forest fires dataset." In: *IEEE Dataport*.

Abid, Faroudja and Nouma Izeboudjen (2019). "Predicting forest fire in algeria using data mining techniques: Case study of the decision tree algorithm." In: *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 363–370.

Achour, S. et al. (2019). "Essais d'optimisation des procédés de clarification et de désinfection d'eaux de barrage de la région de Khenchela (Est algérien)." In: *Larhyss Journal* 37, pp. 151–174.

Alkhatib, Ramez et al. (July 2023). "A brief review of machine learning algorithms in forest fires science." In: *Applied Sciences* 13.14, p. 8275.

Aziz, Abdul, Md Masum Suzon, and Rakib Hasan (Jan. 2025). "A Fuzzy Logic-Based Risk Evaluation and Precaution Level Estimation of Explosive, Flammable, and Toxic Chemicals for Preventing Damages." In: *Heliyon* 11.1.

Balson, Joshua et al. (June 2021). "Identification and characterization of forest fire risk zones leveraging machine learning methods." In: *SMU Data Science Review* 5.2, p. 3.

Barnes, Paul W et al. (Mar. 2022). "Environmental effects of stratospheric ozone depletion, UV radiation, and interactions with climate change: UNEP Environmental Effects Assessment Panel, Update 2021." In: *Photochemical & Photobiological Sciences* 21.3, pp. 275–301.

Belongia, K., T. S. Hogue, and S. Gallo (2023). "Building water resilience in the face of cascading wildfire risks." In: *Science of The Total Environment* 904, p. 166812.

Benbakkar, Hadj Ali et al. (Dec. 2024). "Spatio-temporal analysis and risk management of forest fires (West Algerian region)." In: *Folia Forestalia Polonica. Series A. Forestry* 66.4.

Benkheira, Ak (2018). "Les feux de forêts en Algérie, analyses et perspectives." In: *Séminaire sur les feux de forêts en Algérie-analyse et perspectives, Algérie*.

Bountzouklis, Christos, Dennis M. Fox, and Elena Di Bernardino (Jan. 2021). "Environmental factors affecting wildfire burned area in South-Eastern France, 1970–2019." In: *Natural Hazards and Earth System Sciences* 21.1, pp. 1–24.

Castelli, Mauro, Leonardo Vanneschi, and Aleš Popovič (Mar. 2015). "Predicting burned areas of forest fires: an artificial intelligence approach." In: *Fire Ecology* 11.1, pp. 106–118.

Cui, Zhibo et al. (July 2025). "Mapping Soil Organic Carbon by Integrating Time-Series Sentinel-2 Data, Environmental Covariates and Multiple Ensemble Models." In: *Sensors* 25.7, p. 2184.

De Florio, Mario et al. (Dec. 2025). "Quantification of total uncertainty in the physics-informed reconstruction of CVSim-6 physiology." In: *Philosophical Transactions A* 383.2292, p. 20240221.

Djabri, Ahmed Djaber et al. (June 2023). "Wildfire vulnerability assessment and mapping using remote sensing, gis and weighted overlay method in the eastern aures in Khenchela, Algeria." In: *Geoadria* 28.2.

Dong, Hao et al. (Aug. 2022). "Wildfire prediction model based on spatial and temporal characteristics: A case study of a wildfire in Portugal's Montesinho Natural Park." In: *Sustainability* 14.16, p. 10107.

Flasse, Stéphane P et al. (2004). "Remote Sensing of Vegetation Fires and Its Contribution to a Fire Management Information System." In: *Wildland Fire Management Handbook for Sub-Sahara Africa*, pp. 158–211.

Gazis, Alexandros et al. (2025). "Comprehensive review: sensor technologies, instrumentation, and signal processing in low-power IoT." In: *Academia Engineering* 2.1.

Gebrechorkos, Solomon H et al. (2023). "Global high-resolution drought indices for 1981–2022." In: *Earth System Science Data Discussions* 2023, pp. 1–28.

George, Mayowa B and Enock Okorno Ayiku (2024). "AI-driven fire risk indices integrating climate, fuel, and terrain for wildfire prediction and management." In: *International Journal of Engineering Technology Research and Management* 8.2, p. 67.

Haque, Ahshanul and Hamdy Soliman (Apr. 2025). "Hierarchical Early Wireless Forest Fire Prediction System Utilizing Virtual Sensors." In: *Electronics* 14.8, p. 1634.

Huyen, Chip (2022). *Designing machine learning systems*.

Illarionova, Svetlana et al. (2025). "Exploration of geo-spatial data and machine learning algorithms for robust wildfire occurrence prediction." In: *Scientific Reports* 15.1, p. 10712.

Jolly, W.M., A.M. Hadlow, and K. Huguet (2014). "De-coupling seasonal changes in water content and dry matter to predict live conifer foliar moisture content." In: *International Journal of Wildland Fire*.

Karami, Peyman, Sajad Tavakoli, and Mina Esmaeili (2025). "Fine-scale satellite-based monitoring of temperature and vegetation cover in microclimates, distribution ranges, and landscape connectivity for Neurergus kaiseri (Kaiser's mountain newt) during the breeding season." In: *Ecological Indicators* 170, p. 113077.

Kejzlar, Vojtech et al. (Jan. 2023). "Black box variational bayesian model averaging." In: *The American Statistician* 77.1, pp. 85–96.

Kursa, Miron B. and Witold R. Rudnicki (2010). "Feature selection with the Boruta package." In: *Journal of Statistical Software* 36.11, pp. 1–13.

Lamri, Mohamed et al. (2024). "Deep Learning Prediction of Wildfire Probability Occurrence Using Meteorological Factors from Khenchela Province Forests." In: *International Conference on Electrical Engineering and Control Applications*. Springer, pp. 110–117.

— (May 2025). "Comparative Analysis of Wildfire Prediction Models Across Climatic Regions: Béjaïa and Sidi Bel Abbes (Algeria) vs. Montesinho (Portugal)." In: *Traitement du Signal* 42.3, p. 1219.

Lemesios, Ioannis and George P Petropoulos (2024). "Vegetation regeneration dynamics of a natural mediterranean ecosystem following a wildfire exploiting the LANDSAT archive, google earth engine and geospatial analysis techniques." In: *Remote Sensing Applications: Society and Environment* 34, p. 101153.

Lewis, P. et al. (2010). *Radiative Transfer Modelling for the cha*. Tech. rep. ESA.

Lindahl, Emil and Maxx Wallberg (2022). *Towards latency-aware control using 5G and Edge-based control architectures*.

Liu, Hui et al. (Apr. 2025). "Advancements in Artificial Intelligence Applications for Forest Fire Prediction." In: *Forests* 16.4, p. 704.

Liu, Y. et al. (Apr. 2025a). "Advancements in Artificial Intelligence Applications for Forest Fire Prediction." In: *Remote Sensing* 17.4, p. 845.

— (Apr. 2025b). "Advancements in Artificial Intelligence Applications for Forest Fire Prediction." In: *Remote Sensing* 17.4, p. 845.

López-Cuesta, Miguel et al. (May 2023). "Improving solar radiation nowcasts by blending data-driven, satellite-images-based and all-sky-imagers-based models using machine learning techniques." In: *Remote Sensing* 15.9, p. 2328.

Łoś, H. et al. (July 2021). "Evaluation of XGBoost and LGBM Performance in Tree Species Classification with Sentinel-2 Data." In: *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5803–5806.

Lundberg, Scott M. and Su-In Lee (Dec. 2017). "A Unified Approach to Interpreting Model Predictions." In: *Advances in Neural Information Processing Systems* 30.

Mansoa, Raúl Hoffrén (2024). "Emerging multiscale and 3D remote sensing systems for improved estimation of fuels in Mediterranean forests." In: *Remote Sensing of Environment* 305, p. 114067.

Mansoor, Sheikh et al. (2025). "Integration of smart sensors and IOT in precision agriculture: trends, challenges and future prospectives." In: *Frontiers in Plant Science* 16, p. 1587869.

Matougui, Zakaria and Mohamed Zouidi (Jan. 2025a). "A temporal perspective on the reliability of wildfire hazard assessment based on machine learning and remote sensing data." In: *Earth Science Informatics* 18.1, p. 19.

— (2025b). "Forest fire hazard assessment using remote sensing data and machine learning, case study of Jijel, Algeria." In: *Geomatics, Natural Hazards and Risk* 16.1, pp. 1–20.

Meddour-Sahar, O. et al. (Dec. 2013). "Analysis of forest fires causes in Algeria using regression modeling." In: *Environmental Management* 52.6, pp. 1530–1544.

Miranda, Ana Isabel et al. (2008). "Forest fires and air quality issues in Southern Europe." In: *Developments in Environmental Science* 8, pp. 209–231.

Namburu, Anupama et al. (Mar. 2023). "Forest fire identification in uav imagery using x-mobilenet." In: *Electronics* 12.3, p. 733.

Nolan, R.H., D.M.J.S. Bowman, H. Clarke, et al. (2021). "What Do the Australian Black Summer Fires Signify for the Global Fire Crisis?" In: *Fire* 4.4, p. 97. DOI: `10.3390/fire4040097`.

Nolan, Rachael H. et al. (Aug. 2014). "Changes in evapotranspiration following wildfire in resprouting eucalypt forests." In: *Ecohydrology* 7.5, pp. 1363–1377.

Pausas, Juli G. and Jon E. Keeley (Oct. 2021). "Wildfires and global change." In: *Frontiers in Ecology and the Environment* 19.7, pp. 387–395.

Purnama, Miftahul I et al. (2024). "Predicting forest fire vulnerability using machine learning approaches in The Mediterranean Region: a case study of Türkiye." In: *IOP Conference Series: Earth and Environmental Science*. Vol. 1315, p. 012056.

Radeloff, Volker C et al. (Mar. 2018). "Rapid growth of the US wildland-urban interface raises wildfire risk." In: *Proceedings of the National Academy of Sciences* 115.13, pp. 3314–3319.

Reis, Manuel JCS (Dec. 2025). "AI-Driven Anomaly Detection for Securing IoT Devices in 5G-Enabled Smart Cities." In: *Electronics* 14.12, p. 2492.

Rudin, Cynthia (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." In: *Nature Machine Intelligence* 1.5, pp. 206–215.

Sahar, O et al. (2020). "Diagnostic des incendies de forêt dans le nord centre algérien: cas des provinces boisées et peuplées de boumerdes, bouira et tizi ouzou." In: *Cinq Continents* 10.22, pp. 282–303.

Sahour, Abdelhakim, Farouk Boumehrez, et al. (2026). "Wildfire Risk Prediction using an IoT Sensing Architecture and Fuzzy-Logic-Driven Inference." In: *2026 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*. IEEE, pp. 1–6. DOI: 10.1109/IC_ASET69920.2026.11502207.

Sahour, Abdelhakim, Mohamed Lamri, et al. (2026). "Enhancing Forest Fire Prediction Through Explainable AI-Driven Feature Selection: A Case Study in Khenchela Province, Algeria." In: *2026 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*. IEEE, pp. 1–6. DOI: 10.1109/IC_ASET69920.2026.11502166.

Salaria, Anshika and Amandeep Singh (Apr. 2025). "A Guide for Selection of Wireless Communication Technology for Effective and Robust Early Forest Fire Detection System." In: *Bulletin of Electrical Engineering and Informatics* 14.2, pp. 1026–1035.

Slater, Louise et al. (Mar. 2022). "Hybrid Forecasting: Using Statistics and Machine Learning to Integrate Predictions from Dynamical Models." In: *Hydrology and Earth System Sciences* 26, pp. 1–35.

TALBI, Sarra (2025). "Urban design to enhance flood risk resilience in Tebessa city, Algeria." PhD thesis. University of Tebessa.

Troyanskaya, Olga et al. (June 2001). "Missing value estimation methods for DNA microarrays." In: *Bioinformatics* 17.6, pp. 520–525.

Vom Brocke, Jan, Alan Hevner, and Alexander Maedche (2020). "Introduction to design science research." In: *Design science research. Cases*, pp. 1–13.

Wang, Honggang et al. (Oct. 2024). "An Adaptive Data Rate Algorithm for Power-Constrained End Devices in Long Range Networks." In: *Mathematics* 12.21, p. 3371.

Wang, Junming et al. (Jan. 2010). "Review of satellite remote sensing use in forest health studies." In: *Open Geography Journal* 3.1, pp. 28–42.

Wooster, Martin J et al. (2013). "Thermal remote sensing of active vegetation fires and biomass burning events." In: *Thermal infrared remote sensing: sensors, methods, applications*, pp. 347–390. DOI: 10.1007/978-94-007-6639-6_17.

Xu, Weifeng et al. (Mar. 2025). "Progress and Limitations in Forest Carbon Stock Estimation Using Remote Sensing Technologies: A Comprehensive Review." In: *Forests* 16.3, p. 449.

Yesmin, Farjana and Nusrat Shirmin (Oct. 2025). "A Fuzzy Logic-Based Framework for Explainable Machine Learning in Big Data Analytics." In: *arXiv preprint arXiv:2510.05120*. URL: https://arxiv.org/abs/2510.05120.

Zaidi, Abdelhamid (July 2023). "Predicting wildfires in Algerian forests using machine learning models." In: *Heliyon* 9.7.

Zeb, Shah et al. (Dec. 2021). "Edge intelligence in softwarized 6G: Deep learning-enabled network traffic predictions." In: *2021 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6.

Zhou, Jiahao et al. (Apr. 2024). "Comparing Accuracy of Wildfire Spread Prediction Models under Different Data Deficiency Conditions." In: *Fire* 7.4, p. 141.